

Boosting access to disease data

A new agreement by stakeholders to improve the sharing of flu data should eventually stimulate research on many infectious diseases. Now to make it work.

The pledge by leaders of the rich G8 countries at their summit in Russia in July to boost the meagre quality of international surveillance of infectious diseases correctly identifies many key needs: “better coordination between the animal and human health communities, building laboratory capacities, and full transparency by all nations in sharing, on a timely basis, virus samples.”

A scheme to end secrecy in the sharing of avian-flu samples and data, announced online last week in *Nature* (see page 981), addresses one aspect of the challenge. The Global Initiative on Sharing Avian Influenza Data (GISAID) is modelled on guidelines for sharing data in large-scale biological research (see www.wellcome.ac.uk/doc_wtd003208.html). It is encouraging that so many leading flu researchers have signed up to its principles, under which countries and scientists would immediately share pre-publication samples and data, provided that all those who seek access abide by rules on intellectual property and the attribution of credit. All data would be released in GenBank and other public databases no later than six months after submission. If it works, it would provide a model for the rapid dissemination of data from outbreaks of future emerging diseases.

GISAID tackles several problems. Countries are often reluctant to share outbreak materials and data, as it could compromise their trade or other national interests. Scientists in affected countries can be loathe to cooperate, as they often see little in return by way of scientific cooperation in building surveillance programmes or fighting the disease. Some researchers hoard data, often for years, for competitive reasons (see *Nature* **441**, 1028 & **440**, 255–256; 2006).

A consensus is emerging, however, that full and rapid sharing is ultimately in the best collective interest of research, surveillance and control. In April, OFFLU, the network on avian influenza of the World Organisation for Animal Health (OIE) and the United Nations' Food and Agriculture Organization (FAO), committed itself to making public material on outbreaks in animals. But it has had mixed success, as it relies on the cooperation of affected countries.

Regarding human cases, the World Health Organization (WHO),

while supportive of wider sharing, has been criticized for timidity in pressing countries to release material. But the WHO, like the FAO and OIE, ultimately answers to its member states. Moreover, its priority is not research but the prompt assessment and control of human cases. It has been understandably reluctant to tamper with its existing system, in which countries make samples and data available to a small group of WHO-affiliated labs on a password-protected database. Although imperfect, at least this approach gives it access to much of the genetic, epidemiological and clinical data it needs.

But the system's limits are alarming. For example, when a cluster of human H5N1 cases occurred in Indonesia in May, the WHO had almost no recent animal sequence data available to help it understand the virus's evolution. Rapid sharing of samples and data with scientists across many disciplines is also key to quickly getting a deeper understanding of the virus, and improved drugs and vaccines.

GISAID's broad endorsement of the goal of prompt sharing from multiple stakeholders, often with conflicting interests, is in itself progress and a tribute to the diplomacy of those involved. Tangible evidence of change has also come from the Indonesian government and the US Centers for Disease Control and Prevention, which both announced in August that they would share all flu genomic data; they should be congratulated on having the courage to change policy.

Agreement on the principles of GISAID is only a beginning, however. Prompt progress in establishing the ground rules for sharing will be essential to build confidence and momentum. Governments need to support laboratory capacities in those countries that need it most, where surveillance is weak. And unless donor countries also provide more funds and technical support to fight the disease in animals, which is the reservoir of human cases, we are likely to have more data to share on avian flu than we would like. ■

“A consensus is emerging that full and rapid sharing is ultimately in the best collective interest of research, surveillance and control.”

Rude palaeoanthropology

Controversies over *Homo floresiensis* reflect a flourishing science.

Debate is something on which science thrives. Active disputes are signs of a discipline in rude health, in which discovery piles on discovery, each new fragment of knowledge questioning the one before, until sufficient findings accumulate to decide the matter one way or the other. Or sometimes, a consensus is reached that none of the original protagonists had thought of. This consensus

is always provisional. In contrast, a field in which everyone blandly agrees with everything is a field in stagnation.

That is why the publication of a paper in the *Proceedings of the National Academy of Sciences* (PNAS), robustly countering the identification of *Homo floresiensis* as anything other than a malformed human pygmy (T. Jacob *et al.* *Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.0605563103; 2006), is to be warmly welcomed. The original discovery of a population of bizarre hominids that lived on the island of Flores in Indonesia until relatively recent times (P. Brown *et al.* *Nature* **431**, 1055–1061; 2004) understandably caused a sensation. The idea of an extinct species of ‘little people’ that coexisted with modern humans gripped the public imagination, and journalists



were quick to seize on the nickname of 'hobbit' to stand in for the less wieldy formal nomenclature. Hobbits, of course, are diminutive creatures that sprang from the imagination of J. R. R. Tolkien, who noted, in his novel *The Hobbit*, that hobbits are more elusive nowadays than they once were.

The latter-day hobbit story was spiced up considerably by the characters of the scientists who discovered it — some of whom publicly and not always politely disagreed with one another about the discovery's significance. Not to mention the lively rebuttals from many academic challengers (including but not exclusively the authors of the *PNAS* paper), who contend that to brand the Flores creature as a distinctive species is to create something as fictional as anything invented by Tolkien. Strong words have been exchanged. Skulduggery has been alleged. Accusations fly. This is ideal fodder for journalists whiling away the summer 'silly season'.

No one should be misled, however — the reported invocation of scientific error does not mean that an error has been committed, let alone any kind of misconduct. Is one entitled to ask whether the 2004 *Nature* paper was 'wrong', then? The answer is, robustly, 'yes'.

Palaeoanthropology — the study of the corporeal remains of human evolution — is a notorious arena for splenetic debate. To take the long view, it would have been surprising had the unearthing of the hobbit not led to strongly worded counterblasts, in which critics contended that the new find is really either a diseased human or an ape. Such controversies erupted in 1856 after the discovery of Neanderthal man (said to be a diseased human), and again in 1925 after the announcement of *Australopithecus africanus* (claimed to be a juvenile ape). The status of these creatures as distinct species, neither human nor ape, is now beyond question.

More recently, similar debates followed the discovery of Toumaï (*Sahelanthropus tchadensis*) and the East African taxon *Kenyanthropus platyops*. In the last two, as with the Indonesian 'hobbits', the debate continues — a sign that more remains to be discovered.

But there are obstacles: for two full summer seasons, no Indonesian or foreign group has dug at the Liang Bua cave on Flores. This seems to be an unhealthy by-product of the scientific controversy. The situation should be resolved so that this particular bit of palaeoanthropology can thrive again. ■

Cheap IVF needed

False perceptions are hindering access to new research on *in vitro* fertilization.

In the nearly three decades since the arrival of the world's first test-tube baby in 1978, some two million children have been born using *in vitro* fertilization (IVF). Today, the technique is more popular than ever — but it remains out of reach for many who want it.

This is particularly the case in sub-Saharan Africa, where up to one-third of couples are infertile, causing untold misery for millions of women. Fertility treatments are available in private clinics in several countries in the region, but are far too expensive for most people. A number of scientists are now working to make IVF much cheaper and more accessible (see page 975), and have already come up with some promising alternatives.

But ideas for such methods cannot change lives while they remain stuck on the drawing board, and these developments have yet to be validated in clinical trials. The lack of resources, money, awareness and political will combine to stop them getting off the ground. If this is to change, the issue of infertility needs to be pushed higher up the international agenda.

Perhaps the biggest obstacle is the widespread, but false, perception that infertility isn't a pressing problem in poor African countries. The region also faces daunting challenges of coping with HIV, tuberculosis and malaria. Many argue, understandably, that fighting these should take priority over infertility, which is not directly life-threatening.

But this argument underestimates the devastating social, economic and personal burden of being childless in many poor societies — a burden that mainly falls on women. They are usually blamed for infertility and can be ostracized and assaulted by their families, even driven to suicide or killed. By supporting the development of low-cost IVF, governments can help make such treatments more widely available.

Another frequently heard argument holds that it is inappropriate to prioritize treating infertility in countries that are already struggling to support their fast-growing populations. But the 1948 UN Universal Declaration of Human Rights states: "Men and women of full age, without any limitation due to race, nationality or religion, have the right to marry and to found a family." In cases where parents are unable to have children, this ought to imply the right to access to fertility treatments.

Too many women in Africa are deprived of the right to an education, freedom from abuse, and access to decent health care and contraception. As well as being desirable in its own right, the education and empowerment of women has been shown to be a critically important factor in slowing the high birth rate that critics of fertility treatments in Africa say they are concerned about.

Raising the status of women and broadening access to contraception, particularly condoms, would also help to control the sexually transmitted diseases that are the main causes of infertility in sub-Saharan Africa. The crippling social taboos surrounding childlessness also need to be challenged. But changing age-old prejudices is going to take a great deal of time and concerted effort by governments and by pressure groups such as Chipso Chedu, a Harare-based trust established by Betty Chishava, a Zimbabwean woman cast out of her home for failing to bear children.

In the meantime, the scientific and medical community can play its part by creatively questioning whether exotic new equipment and state-of-the-art drug therapies are necessary for safe and effective fertility treatments. Some scientists argue that much of this is dispensable and is largely driven by the affluent end of the fertility market. Indeed, it would be a nice twist if low-cost fertility treatments designed to help impoverished African couples could one day make IVF affordable for less-well-off couples in rich countries too. ■

"Women are usually blamed for infertility, and can be ostracized and assaulted by their families."

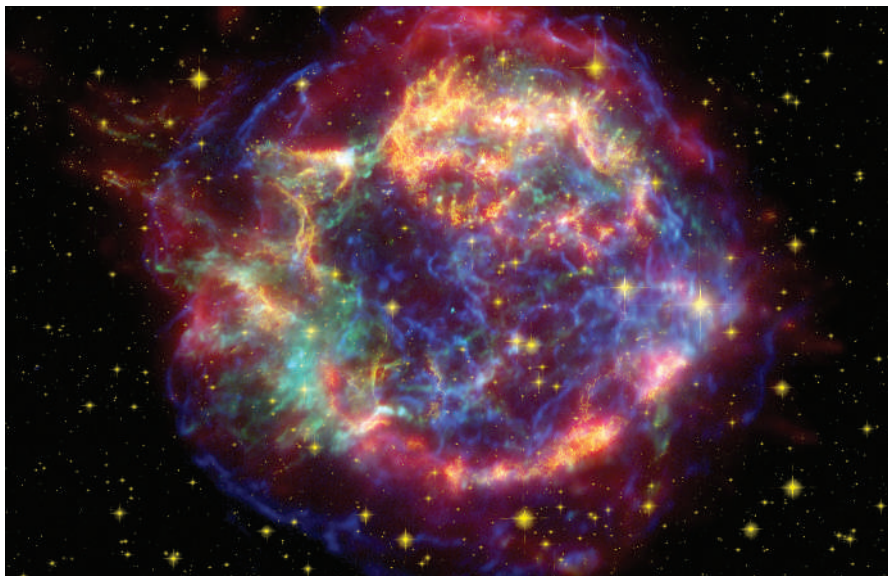
RESEARCH HIGHLIGHTS

A shot in the dark

Nature Phys. doi:10.1038/nphys391 (2006)

Cosmic rays have long been thought to be accelerated to high speeds by exploding stars known as supernovae. But direct evidence has been difficult to come by. Unlike photons, which travel directly from the source, cosmic rays' paths can be twisted by intervening magnetic fields.

Now Michael Stage of the University of Massachusetts in Amherst and his colleagues have caught cosmic rays in the making. Using the Chandra X-ray Observatory, the team identified regions of the Cassiopeia A supernova remnant (pictured) where electrons were being accelerated by the supernova's shockwave. Some electrons were accelerated almost as fast as theoretically possible, and their overall spectrum matched well with cosmic-ray theory.



NASA/JPL-CALTECH/O. KRAUSE (STEWART OBSERVATORY)

CHEMICAL BIOLOGY

Changing the code

Nature Methods 3, 729-735 (2006)

Can DNA be improved on? Although it is prodigious at encoding genetic information, it does so with an alphabet of only four letters. If this alphabet were extended, more information could be stored.

Ichiro Hirao and Shigeyuki Yokoyama at the RIKEN Genomic Sciences Center in Yokohama, Japan, and their colleagues have prepared unnatural DNA bases that can be incorporated into DNA. Unlike natural bases, which use hydrogen bonds to pair up across the DNA duplex, these unnatural bases form pairs using hydrophobic interactions.

The team showed that standard enzymes, when exposed to the appropriate molecular building blocks, can replicate DNA containing these unnatural base pairs. Such DNA even acts as a template to incorporate unnatural bases into RNA, through translation.

PHYSICS

Muscle power

Opt. Lett. 31, 2610-2612 (2006)

Video monitors of the future might use artificial muscles to get their colours exactly the right shades.

Manuel Aschwendt and Andreas Stemmer of ETH Zurich, Switzerland, have created a variation of a light-spreading device known as a diffraction grating that includes a membrane-like polymer that deforms when a voltage is applied to it. Changing the voltage

adjusts the position of the membrane — the artificial muscle — which changes the colour of the diffracted light.

Getting a single pixel to change from blue to red took thousands of volts, but the team hopes to lower the required voltage to commercially viable levels.

CLIMATOLOGY

Methane blast

Science 313, 1109-1112 (2006)

The reasons for the marked rise in atmospheric methane concentrations at the end of the last ice age remain controversial.

Hinrich Schaefer of the University of Victoria, Canada, and his colleagues looked at methane trapped in western Greenland ice samples (pictured below) spanning an 800-year period of warming about 12,000 years ago. Compared with present-day atmospheric methane, the gas was less depleted in the heavier carbon isotope carbon-13 than expected. That



suggests a different group of sources from today was putting methane into the atmosphere back then.

Emissions from tropical wetlands, or from living plants — a recently discovered source — may account for the discrepancy, the authors say. The fact that the ratio of the isotopes stays the same throughout seems to rule out a burst of methane from clathrates in the sea bed.

BIOCHEMISTRY

Cold comfort

Curr. Biol. 16, 1591-1605 (2006)

Nerve-endings in the skin that respond to cold temperatures and to chemicals that mimic the cold, such as menthol, could represent the coolest strategy yet for treating chronic pain due to nerve injury. These nerves send signals to the spinal cord that block pain messages from elsewhere in the body.

Susan Fleetwood-Walker of the University of Edinburgh, UK, and her colleagues induced chronic pain in rats, then treated them with either menthol or icilin, another cooling chemical. Both treatments activated the 'coolness' receptor TRPM8 in the skin and had an analgesic effect.

MOLECULAR BIOLOGY

A question of control

Nature Struct. Mol. Biol. doi:10.1038/nsmb1138 (2006)

Current methods used to predict which genes are controlled by tiny microRNAs (miRNAs) could be inaccurate, say Dominic Didiano and Oliver Hobert at Columbia University Medical Center in New York.

E. BROOK

Genes controlled by miRNAs are typically found by bioinformatic programs, which scour genomic data for short sequences that pair up with a core region of the miRNA molecule. The researchers tested 13 such candidate sites in the worm *Caenorhabditis elegans* and found that none are controlled by the putative miRNA.

They show that unknown features outside an miRNA's binding site seem to determine whether it controls a gene, and warn that many of the thousands of genes thought to be controlled by miRNAs may need to be re-evaluated.

BIOLOGY

Jaws in a flash

Proc. Natl Acad. Sci. USA **103**, 12787–12792 (2006)
Super-speedy jaws help an ant of Central and South America in several ways, biologists have found.

High-speed imaging has shown that the trap-jaw ant, *Odontomachus bauri* (pictured right), can snap its jaws shut at a speed of up to 64 metres per second. The ant can use this motion to stun prey, knock away an enemy, or even fling itself into the air (see news@nature for video at <http://tinyurl.com/q2v1r>).

A team led by Sheila Patek at the University of California in Berkeley and Andrew Suarez at the University of Illinois at Urbana–Champaign says the many uses of the jaw suggests that 'extreme animal movements' can have a number of evolutionary advantages.

BIOCHEMISTRY

Gut instinct

Science **313**, 1126–1130 (2006)
In the bacterial battle for intestinal territory, boundaries are strictly enforced. If the balance is thrown off kilter, this can trigger unpleasant disorders such as inflammatory bowel disease.

Researchers have now identified an antimicrobial protein that may help maintain order. Benign bacteria living in the guts of mice stimulate production of the protein, which binds to sugars protruding from some bacteria, say Lora Hooper and her colleagues at the University of Texas Southwestern Medical Center at Dallas. The mouse protein and its human counterpart rupture the cell walls of unwanted bacteria.

Mouse pups express the gene encoding this protein at particularly high levels after weaning, when intestinal flora change and influxes of antibodies from their mother cease. This suggests that the protein has an innate role in intestinal defence.



A. WILD/MYRMECOS.NET

PHYSICS

A soft touch

Appl. Phys. Lett. **89**, 073501 (2006)
Modifying a kind of electronic transistor could lead to the development of new pressure sensors for prosthetic skin.

Ingrid Graz, based at Johannes Kepler University in Linz, Austria, and her colleagues created the sensors by placing a film of polypropylene ferroelectret, a non-polar material that retains electric charge, on top of a field-effect transistor (FET). The device can detect pressure applied to the film, flipping on or off in response to touch.

The authors suggest that layered arrays of FETs and pressure-sensitive ferroelectrets could be used to create 'sensor skins' for use in prosthetic or wearable electronic devices.

EVOLUTIONARY BIOLOGY

Made for each other

PLoS Biol. **4**, e235 (2006)
Parasites, like lovers, are very particular about their relationships. Nicole Gerardo of the University of Arizona and her colleagues report on the specificities that keep parasitic fungi of the genus *Escovopsis* loyal to their prey, the fungi cultivated by certain ant species.

The parasites are chemically attracted to their hosts in a way in which they are not to other species. The hosts produce antibiotics that, although tolerable to their accustomed parasites, inhibit other closely related species. Only very occasionally are the cultivated fungi susceptible to parasites that they have not previously encountered, making it hard for the parasites to switch hosts.

JOURNAL CLUB

Cole Miller
University of Maryland, College Park, Maryland, USA

Fast-moving stars give this astronomer new reasons to admire the heavens.

I grew up in the countryside, and would often look up at the night sky. Like billions before me, I concluded that the stars are essentially immobile. Indeed, in our vicinity, a typical star moves at only 20 kilometres per second relative to the average motion of nearby stars.

Recently, however, some exceptions to this slow grandeur have been discovered. Warren Brown of the Smithsonian Astrophysical Observatory in Cambridge, Massachusetts, and his colleagues (W. R. Brown *et al.* *Astrophys. J.* **640**, L35–L38; 2006) report observations of so-called 'hypervelocity' stars. These stars, tens of thousands of light years away, have speeds of 500–700 kilometres per second, which is sufficient to escape our Galaxy entirely.

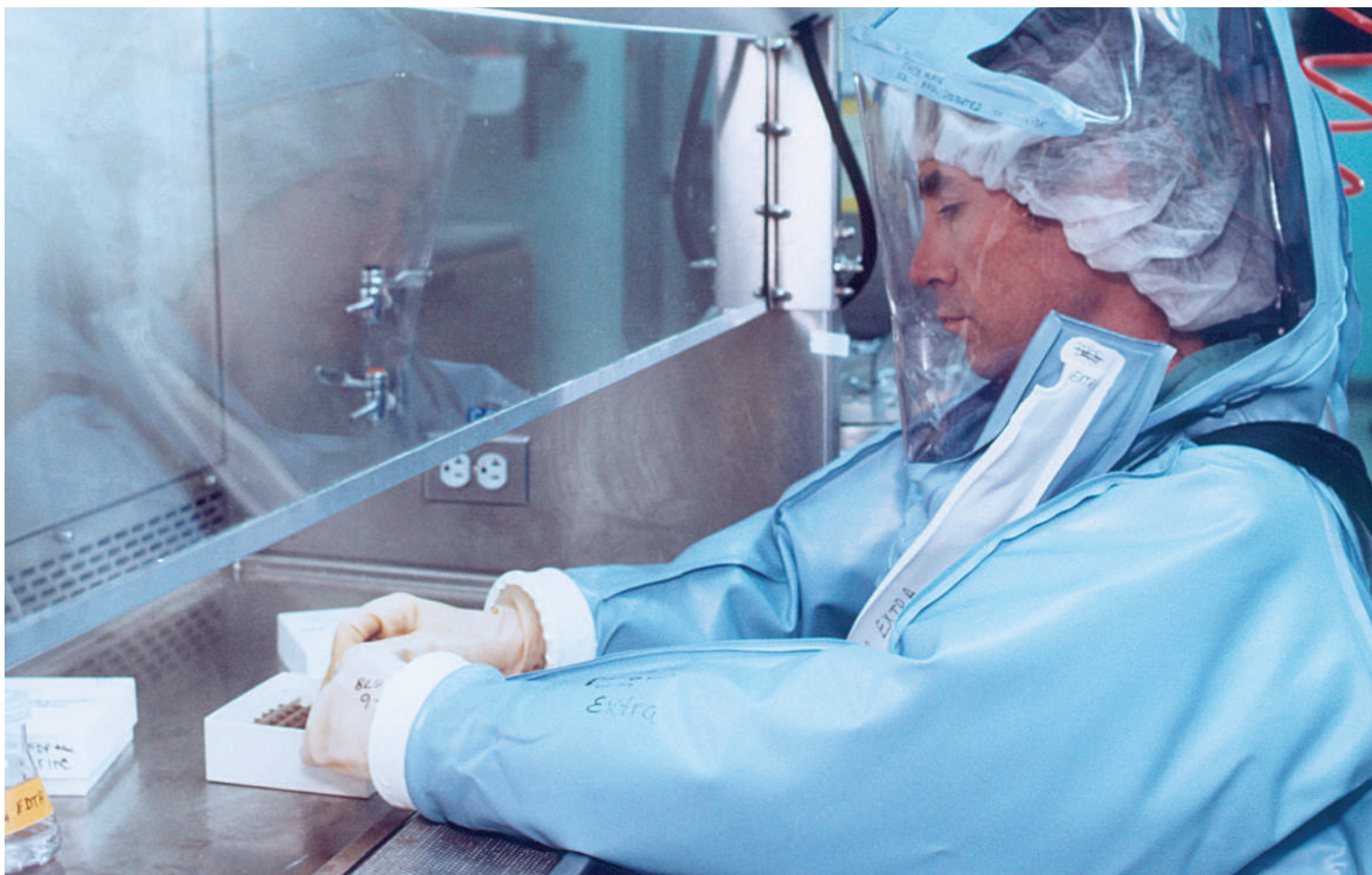
Remarkably, the best explanation for the behaviour of these objects was proposed before their discovery (J. G. Hills *Nature* **331**, 687–689; 1988). If a closely bound pair of stars plunges near a supermassive black hole lurking at the galaxy's centre, purely gravitational effects can separate the stellar binary, leaving one star in orbit around the black hole and sending the other on a one-way trip to intergalactic space. A Galactic-Centre origin for these stars is consistent with their current ages, positions and projected velocities.

Even more dramatic ejections could happen when stars interact with the binary supermassive black holes inside recently merged galaxies.

I live in the city now, so the skies are not as favourable for perusal. When I do look up, though, I have a newfound respect for the dynamism of the heavens and the role of black holes in the cosmic dance.

NEWS

CDC



Locals rally to combat biodefence labs

As the US government picks up the pace of building high-security biodefence laboratories, community groups and watchdogs are ramping up their protests.

The latest clash centres on Fort Detrick, an army facility in Frederick, Maryland, that has long been home to biosecurity labs. The government is planning to overhaul the existing facilities and build a new biodefence research complex. Construction has already begun on one component: the Department of Homeland Security's National Biodefence Analysis and Countermeasures Center (NBACC). This is slated for completion in June 2008.

But critics want to halt work on this facility and others in the works nationwide. On 30 August, opponents were scheduled to air their concerns at a public meeting in Frederick about Fort Detrick's expansion.

"From almost any way of looking at it, this makes absolutely no sense," says Barry Kissin, a lawyer and congressional candidate from

Frederick, about the planned facility. "This does the opposite of provide for our security, at great expense and great hazard."

Although local communities often protest about biodefence labs in their midst, the \$105-million NBACC does stand out. Plans inadvertently posted on the Internet suggest that personnel there will conduct exercises known as 'red-teaming'. These would involve creating and testing biothreat agents thought to belong to enemy arsenals. The Department of Homeland Security is also attempting to classify the facility, which means activities there would be off-limits to public enquiry.

The NBACC will include labs operating at the highest biosecurity level — biosafety level 4 — which handle the deadliest pathogens. This would be in addition to several biosafety

level 4 facilities already in existence at the Fort Detrick campus. But critics say more labs will increase the threat to the surrounding community — they say pathogens could escape or be removed surreptitiously from the labs. Anthrax that was used in the unsolved mail attacks of late 2001, for instance, is thought to have come from a research lab.

Opponents also charge that the facility runs the risk of spurring other countries to ramp up biowarfare activities: unless inspectors are allowed to investigate the site, some might suspect the United States of creating offensive weapons.

"It's a really big mistake to classify the entire NBACC facil-

ity," says Alan Pearson, an expert on biological weapons at the Center for Arms Control and Non-proliferation in Washington DC. "Clearly there are going to be some aspects of the work

"How will the United States assure the world that its research is defensive if it's being conducted at a classified nuclear-weapons laboratory?"

**SHUTTLE LAUNCH**

For updates on Atlantis, read our newsblog coverage online.

www.nature.com/news

T. CRYDER/NASA

Danger signals: the number of US labs that can handle deadly pathogens is rising.

that ought to be classified, but those ought to be minimal. Otherwise you start generating suspicions and, frankly, generating excuses for other countries."

The department says that classification is necessary to prevent other nations from obtaining information about US weak points. "Providing a secure environment for the handling of sensitive information in this way makes sense, and will not allow our enemies to gain the advantage should vulnerabilities be revealed," says Christopher Kelly, a spokesman for the department.

Critics are also worried that the Department of Homeland Security could attempt to classify another project it has in the works: a \$450-million complex of high-biosecurity labs and testing grounds called the National Bio and Agro-Defense Facility. The department is holding a nationwide competition to determine where the lab will be located; 12 sites remain in the running. Kelly says the department has no plans to classify the complex in question. But one site that has made the shortlist is the Lawrence Livermore National Laboratory in California — which is already a classified facility.

"How will the United States assure the rest of the world that the research is completely defensive if it's being conducted at a classified nuclear-weapons laboratory?" asks Marylia Kelley, executive director of Tri-Valley CAREs, a Livermore-based group that monitors the national laboratory.

NBACC and the bio-agro facility are just part of a recent boom in biodefence spending in the United States. The federal government has spent \$36 billion to combat bioterrorism since the terrorist attacks of 11 September 2001, according to an analysis by the Center for Arms Control and Non-proliferation. Three additional biosafety level 4 labs are in the works, including one at Boston University in Massachusetts that has been heavily protested. Fourteen new biosecurity level 2 and 3 labs are also planned.

Pearson argues that the country should spend more on working to prevent bioterrorism in the first place, by strengthening the United Nations' biological and chemical weapons conventions, and improving supervision of its own research. "It wouldn't take much money to strengthen prevention," he says, "and it would do a lot more to keep us safe."

Erika Check

Plan to pool bird-flu data takes off

A bid by leading researchers to bring into the open data on bird flu that were previously kept behind closed doors has met with cautious optimism from observers.

Some 70 avian-flu scientists from all corners of the globe have signed up to the Global Initiative on Sharing Avian Influenza Data (GISAID). A letter outlining the agreement was published online last week in *Nature* and appears in this week's issue (see page 981).

The move aims to resolve issues that have seen some countries and organizations come under fire for hoarding genetic information about avian flu strains. The reasons for their reluctance to share are varied: some, for example, fear that others might use the data without properly crediting the researchers involved. And there were concerns over whether countries worst hit by bird flu would benefit from the drugs and vaccines developed from the sequences they provide.

Precise details of the GISAID agreement are still being thrashed out. But the idea is that participants will place genetic data into secure sections of existing online databases as soon as possible after producing and analysing them.

Peter Palese, who studies flu

viruses at Mount Sinai School of Medicine in New York, describes GISAID as "a very positive development". There are signatories from many of the countries affected by bird flu, including China, Indonesia, Thailand and Vietnam. And the Indonesian government has thrown its support behind free access to data. But Palese questions whether the agreement will work without the full involvement of all the scientists and governments in these countries.

Menno de Jong, an avian-flu researcher at the Hospital for Tropical Diseases in Ho Chi Minh City, Vietnam, also welcomes the initiative. But he cautions that care should be taken to ensure GISAID doesn't become "another, albeit bigger, old boys' network". He suggests that increasing the capacity of poor countries affected by the disease to do their own research, enabling them to be more equal partners in collaborations, might be an extra encouragement for them to share their data.

GISAID proposes pooling its data using the International Nucleotide Sequence Database Collaboration, a network of three major public databases. The data will initially be accessible only to researchers

who have signed up to GISAID, but the information will become open to the public no more than six months after it is deposited.

GISAID members have also agreed to collaborate with, and appropriately credit, all relevant researchers in any resulting publications and intellectual-property agreements.

Clinical and epidemiological data are included in the agreement. The hope is that researchers will be able to compare new strains against others quickly, for example to track whether a virus is acquiring mutations or becoming resistant to drugs.

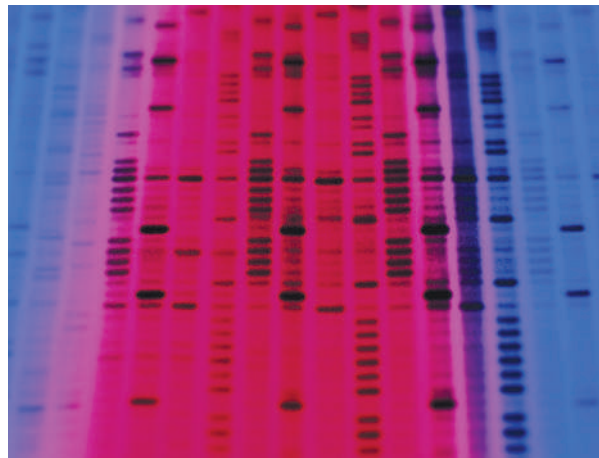
Until now, access to many genetic sequences has been restricted to a global network of flu labs associated with the World Health Organization (WHO). Dick Thompson, WHO spokesman in Geneva, says that the organization supports the sharing of sequence data. He notes that the WHO has sometimes been portrayed as wanting to keep data secret but says that is not the case. "Sometimes countries have legitimate reservations and we have to work with that," he says.

John Sulston, of the Wellcome Trust Sanger Institute in Cambridge, UK, says that he supports GISAID as a compromise. But he adds that because bird flu is spreading around the globe so fast, "I still think it's a good idea to release data immediately."

But virologist Ilaria Capua at the Experimental Animal Healthcare Institute in Padua, Italy, who campaigned for the initiative with Hollywood media adviser Peter Bogner, says she is very happy with the result. She adds that the framework could be used for other emerging infectious diseases. "If a new SARS knocks on our door, we have a system in place."

Helen Pearson

K. CURTIS/SPL



Lining up: access to DNA sequence data is vital for fighting bird flu.



C. & A. PURCELL/CORBIS

ZOO NEWS

A fishy story

The first fish ever to receive chemotherapy for cancer has died, probably of old age. Bubba, a large grouper, also changed gender in the mid-1990s — but that's quite common for a fish.

NUMBER CRUNCH

Last week four mathematicians were awarded the Fields Medal, the mathematical equivalent of the Nobel prize.

27,000 Google hits exist for Andrei Okounkov — recognized for his work on representation theory and randomness.

50,800 is the number of hits for Wendelin Werner — praised for geometric insights into statistical processes and field theory.

71,400 hits exist for Terence Tao — harmonic analysis and number theory.

516,000 hits exist for Grigory Perelman — nominated for his work on solving the Poincaré conjecture. He declined the prize because he didn't want the publicity.

SCORECARD

Ethical stem cells
Scientists may still be debating whether a paper published in *Nature* last week by Robert Lanza and his colleagues at stem-cell company ACT represents an ethically acceptable source of human embryonic stem cells. But the business world seems convinced — the company's share price leapt from \$0.45 to \$2.30 in just 10 hours after the news became public.

Breast cancer research
"Breast cancer has become more about making money for corporate sponsors than funding innovative treatments," according to Samantha King of Queen's University in Ontario. Strong words, but perhaps not unjustified. When she studied one high-profile firm's walk for breast cancer, she found that only 64% of the money raised went to breast cancer organizations.

Source: Pink Ribbons (Univ. Minnesota Press).

Cash-strapped research ship must earn its living

TOKYO

The Japanese ship *Chikyu* is the most sophisticated research vessel ever, and scientists are keen to start using it. But before it can begin probing seismic zones of the sea floor, a lack of funds means it must spend months looking for oil.

Chikyu has run aground on high oil prices, which have more than doubled since the US\$30-a-barrel days when much of the planning was done. This raises the cost of operating the ship, which consumes 110,000 litres of oil per day even when it's stationary and drilling. It also increases the cost of hiring crew, as *Chikyu* is competing with an oil industry scrambling to ramp up production to take advantage of high prices. As a result, estimates of annual running costs have jumped by 50%.

The vessel, which cost ¥60 billion (\$500 million), was built to take the leading role in the Integrated Ocean Drilling Project (IODP), due to start in September 2007. The IODP is a collaboration of Japanese, US and European scientists who study rock and ice-core samples in search of clues about Earth's evolution. Japanese researchers are especially interested in collecting samples from deep fault regions, which can help show how tectonic plates shift during earthquakes.

This month, *Chikyu* headed off to practice core sampling and analysis off the Japanese coast, after which it was due to join the IODP. But the rising running costs caused the ship's managers at the Japan Agency for Marine-Earth Science and Technology (JAMSTEC) to change plans. From November the vessel will spend nine months working in the Indian Ocean off Kenya for an oil exploration project run by Australian petroleum company Woodside. "The alternative was just to have it sit in port," says JAMSTEC director Kiyoshi Suyehiro.

As a non-profit organization, JAMSTEC won't receive any money from Woodside. But drilling for oil will at least be an opportunity to train the crew and test the riser drilling system, which is common in oil drilling but hasn't yet been used in research. Riser drilling uses circulating mud to prevent the hole from collapsing under high pressures, and so allows deeper digging.

Chikyu is designed to reach down 7,000 metres, even in deep waters — the only research vessel that can drill so deep. The oil project in



JAMSTEC/IODP/MI

The state-of-the-art research drilling vessel *Chikyu* is being kept afloat by exploring for oil.

Kenya will involve drilling 2-kilometre wells in ocean up to 3 kilometres deep, so it will provide a reasonable trial. The rough currents will also test *Chikyu*'s thruster system, which keeps its position stable.

But much of *Chikyu*'s research equipment will sit idle. For example, as the core samples will belong to Woodside, the crew won't be able to analyse them using the vessel's magnetic resonance imaging devices.

Chikyu is still on course to begin scientific drilling in September 2007, but high running costs are likely to continue to plague the ship. Its first project will be to dig into the Nankai Trough, an earthquake-generating zone off the Japanese coast. The first stage, drilling down to 6,000 metres, is scheduled to take five years, but is likely to take much longer than expected unless oil prices drop or extra funding is secured to allow the vessel to operate continuously.

JAMSTEC, which refuses to shift funds from its other ocean exploration projects, is pressing for extra money from the Japanese government. It has convinced the science ministry but will have a harder time persuading the finance ministry, which is trying to rein in the Japanese budget by ¥14.3 trillion over the next five years.

David Cyranoski



NATURE'S PODCAST
Download Nature's free
audio news show.
[www.nature.com/
nature/podcast](http://www.nature.com/nature/podcast)



The eyes have it: astronomers have criticized the vote to define a planet.

M. CIZEK/AFP/GETTY

Pluto: the backlash begins

The future of the Solar System — or at least that of some of its nomenclature — may be thrown into turmoil by scientists who are calling for a boycott of a new definition of a planet.

On 24 August, delegates at the general assembly of the International Astronomical Union (IAU) in Prague voted to define the planets of the Solar System by three criteria. To deserve planet status, the assembly agreed, a chunk of rock or ball of gas must be big enough for its gravity to have made it round, must orbit the Sun but not be a satellite of another planet, and must have cleared other bits of debris from its orbit. Round objects, including Pluto, that failed on the final count became not planets but 'dwarf planets'.

The definition originally proposed on 16 August by the IAU would have had just two criteria — roundness and not being a moon. This was rejected by members at the meeting where the three-part definition was voted on as the final word on the subject. But many IAU members were not in Prague for the vote, and some are furious at the outcome.

"I am just disgusted by the way the IAU, which is meant to represent the best in science, handled this matter," says Alan Stern, a planetary scientist at the Southwest Research Institute in Boulder, Colorado. As principal

investigator for NASA's New Horizons mission to Pluto, he has a particular interest in its status. But he says the issue is not really Pluto's status so much as the idea of putting objects in orbital contexts. "We do not classify objects in astronomy by what they are near," he says. "We classify them by their properties."

The day after the new definition was unveiled, Stern was among a dozen scientists who launched a petition to contest it. By e-mail, they sought the support of their colleagues for the following statement: "We, as planetary scientists and astronomers, do not agree with the IAU's definition of a planet, nor will we use it. A better definition is needed." More than

200 people had added their names to the petition as *Nature* went to press on Tuesday.

Stern thinks that requiring a planet to have 'cleared its orbit' rules out some of the Solar System's other eight planets. These include Neptune, the orbit of which is crossed by Pluto, and Jupiter, which shares its orbit round the Sun with the Trojan asteroids.

The 'clearing' criterion was introduced when astronomers who study the dynamics of the Solar System insisted that the definition should recognize their idea of what constitutes a planet — an object with a mass that dominates its orbital zone. Owen Gingerich, chair of the committee that proposed the 16 August

resolution, thinks the IAU had no choice but to bend to the dynamicists' demands. "They may not have had a majority for anything positive, but they could rouse a strong negative majority simply because there are so many little fiefdoms," he explains.

"The dynamics part of the definition is a rather complex one," says Ron Ekers, past IAU president. Couching the idea in terms of a planet 'clearing its orbit' was intended to make the issue easier for the public to understand. But it may well end up confusing matters.

Some organizations have already said they will accept the IAU's new definition. *Encyclopaedia Britannica*, for example, issued a statement saying that some of its articles on Pluto and the Solar System were updated online the same day the IAU's pronouncement was made. According to a spokesperson, later revisions may reflect any uncertainty, but "the vote by the IAU is considered binding — until the next vote, whether it's next year or next century". NASA, too, promised to abide by the definition, adding that it will "continue pursuing exploration of the most scientifically interesting objects in the solar system, regardless of how they are categorized".

Others are waiting to see how strong the counter-movement becomes. A black ribbon was tied around the Pluto panel at the Smithsonian Institution's National Air and Space Museum in Washington DC when the IAU's verdict was announced, says curator Andrew Johnston. But

**"We do not classify
objects in astronomy
by what they are near,
we classify them by
their properties."**

—Alan Stern

it has since been removed. "We're going to let things calm down for at least a few weeks before we decide to make any changes," he says.

One thing that particularly irks critics is the way the decision was made. The IAU has nearly 9,000 members, but only 2,500 people attended the Prague meeting and only a few hundred were present for the vote. The IAU should have used the Internet to gauge wider opinion, and then allowed electronic voting, according to those who oppose the definition.

"The IAU seems to be rooted in the pre-Internet age," says Mark Sykes, director of the Planetary Science Institute in Tucson, Arizona, who instigated the petition. "The rules of the IAU say that resolutions are passed by those present and voting," says Catherine Cesarsky, director of the European Southern Observatory and newly elected president of the IAU.

Sykes admits that a "better definition" might be hard to come by, but is still pressing for the current one to be scrapped. He thinks the IAU would be better off without any definition at all rather than the one they have chosen. "If they can determine that this process was flawed and nullify it, then I think that would be in their best interests," he says.

"If enough people are completely unhappy, we could go through the process again," says Ekers. But a new resolution would have to wait for the next general assembly in 2009 in Rio de Janeiro. The IAU may issue a clarifying statement in the next week or two, but is hesitating to do so now. "Perhaps we need to make our next statement when things are a little less emotional," says Ekers. ■

Jenny Hogan

Diary of a planet's demise

While attending the International Astronomical Union's meeting in Prague, **Jenny Hogan** kept the world up to date on the Pluto debate through our newsblog. Edited excerpts:

Monday 21 August

The proposal to define a planet as anything round that isn't a moon, and thus increase the tally in our Solar System to 12, is scheduled for discussion at lunchtime tomorrow. But many astronomers have already conveyed their objections to the executive committee of the International Astronomical Union (IAU) by e-mail — and some are supporting a second, rival definition.

This alternative definition argues that a planet, as well as being round, must also be "by far the largest object in its local population". This definition knocks Pluto off its planetary pedestal (although it offers it concessionary 'dwarf planet' status), and destroys the chances of promotion for Ceres, queen of the asteroid belt.

Of the 100 people in the closed meeting last Friday where the alternative definition was floated, a show of hands showed about 50 for it and only 20 for the IAU's suggestion.

23:00 My dinner companions tonight include some (very tired) members of the Planet Definition Committee. They say they have received hundreds of e-mails over the past few days from geologists complaining

about the proposal in the original definition to use 'pluton' to mean an object in the same class as Pluto. Pluton is a term of long-standing and wide use in geology, where it refers to an intrusion of igneous rock.

Another problem has emerged in translation. The French name for Pluto is — you've guessed it — Pluton. The definition committee thought this linguistic borrowing would give the pluton label special appeal for French-speaking astronomers, but apparently some of them object.

All this leads to speculation that tomorrow's revised definition, whatever other changes it contains, will include a replacement word for 'pluton'.

Tuesday 22 August

15:00 For people who often tell journalists that defining a planet is a meaningless labeling exercise, astronomers actually seem to care a great deal. The open discussion on what makes a planet stopped just short of fisticuffs.

The official resolution has been divided into three parts, each of which will be voted on separately on Thursday at the closing ceremony. These cover the requirement of roundness; the distinction between a binary planet

Dwarf planet in quotes

"I'm here. I'm a sphere. Get used to it."

Pluto itself, talking to Gady Epstein of the *Baltimore Sun* about its recent demotion.

"I don't know about the public, but... the astrologers will be upset."

Patrick Moore, astronomer and veteran presenter of the BBC's *The Sky at Night*.

"Please don't turn Pluto into a dwarf planet because that makes me sad. I'll miss Pluto a lot."

Daniel Dauber, aged six, on *Nature's* Newsblog.

"This is as if botanists had

found something between trees and bushes and invented the word 'animal' to describe it."

Allen Glazner of the University of North Carolina, Chapel Hill, on the proposal to call dwarf planets 'plutons' — a term that geologists have long used to describe certain bodies of rock.

"Since the term is not in the Microsoft Word or WordPerfect spellcheckers, we thought it was not that common."

Owen Gingerich, chairman of the Planet Definition Committee, which proposed the use of the term pluton.

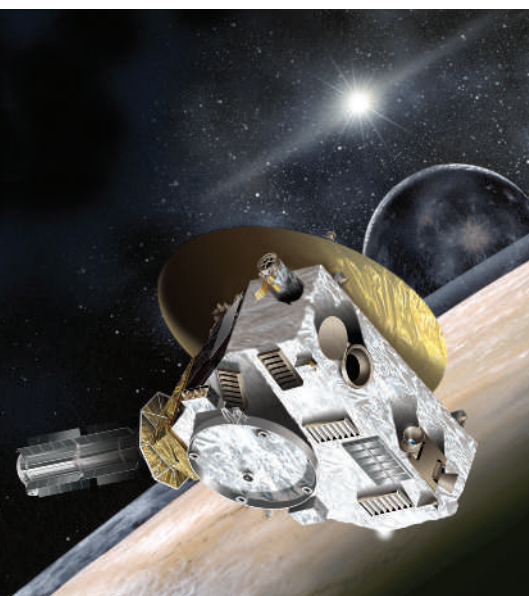
"The comments were intelligent, but they came with a passion that makes me think this debate has a non-intelligent dimension."

Paul Murdin, Cambridge astronomer, at the annual meeting of the International Astronomical Union.

"It's over, it's done."

Richard Binzel, an astronomer at the Massachusetts Institute of Technology and member of the Planet Definition Committee, on the Prague vote.

Sources: *Baltimore Sun*, *Guardian*, *Nature*, news@nature.



NASA's New Horizons probe should reach Pluto in 2015, regardless of whether it is still a planet.

**CHEMISTRY BLOG**

Live reports from the first European Chemistry Congress in Budapest.

<http://blogs.nature.com/news>

L. HOLM NIELSEN/IAU

and a planet-moon system; and the naming of Pluto-like objects.

Within seconds of comments being invited, queues form at the microphones. One by one, astronomers denounce the definition in tones ranging from offended to furious. The representatives of the Planet Definition Committee slump into their chairs, heads propped on their hands.

Andrea Milani of the University of Pisa is first to reach a microphone. He articulates the concerns of the 'dynamicists' — astronomers interested in orbits, many of whom feel strongly that the condition of dominating an orbital zone should be a central part of the definition. Milani becomes more incensed as he speaks, ending by saying "your paper is a kind of offence to the entire dynamical community".

Meanwhile, those who work on extrasolar planets — some with many times the mass of Jupiter — feel that their field has been neglected. Why does the definition not set an upper mass limit? As this point was raised again and again, IAU president Ron Ekers became more and more frustrated. "We want your input, but not right now," he eventually snapped.

17:30 We are now on version three of the planet definition. I was expecting another lively show of dissent — but it is not to be, thanks to Jocelyn Bell Burnell, the astronomer who discovered the first pulsar. A member of the IAU's resolution committee, which decides what gets voted on, she takes formidable control of the meeting. With only 45 minutes available, she requires comments to be no more than 'elevator pitches' — sold in the time it takes a lift to travel one floor. The astronomers meekly follow her orders.

The latest version requires that a planet be both round and, at the insistence of the dynamicists, dominant. Round objects that don't dominate their local orbital zone are 'dwarf planets'. Bell Burnell spells out the consequences: "This means that Pluto is a dwarf planet, but it is not a planet." Would that be acceptable to the assembled astronomers?

It seems so. In a quick show of hands, more arms are raised in favour than against.

Thursday 24 August

The final text of the resolution (version four by my count) is posted in today's edition of the conference newspaper *Nuncio Sidereo III*. According to this resolution, the Solar System has eight top-flight planets, with Pluto in a second class of dwarf planets. Separate votes will be held on whether to label these top-flight planets 'classical planets' and what, if anything, to do about



Moment of truth: Jocelyn Bell Burnell uses props to liven up the voting on whether Pluto is a planet.

putting Pluto and other round trans-neptunian snowballs into a 'plutonian object' category. "Only minor corrections can be accommodated at this stage," the paper warns.

11:30 I'm skipping down the stairs of the conference centre on my way to a 10.30 interview (not about planets) when I encounter a charge of scientists led by the esteemed Brian Marsden. "You're the press," one of his cohort notices. "Show us to the press room."

I retrace my steps. Marsden has, for many years, been the head of the Minor Planet Center at Harvard, a clearing house for orbital data on asteroids and comets. (This week's redefinitions are set to turn them into 'small Solar-System bodies'.) Today marks his retirement, but he enters the press room with youthful vigour.

He holds up an A4 sheet of paper, on which is written in very large letters the word 'Planetino'. "Planetino is what they say in the resolution is a dwarf planet," he proclaims.

Pointing to the ten or so astronomers straggling in behind him, Marsden says his proposal to call 'dwarf planets' 'planetinos' instead has support from representatives of Uruguay, Brazil, the Czech Republic, the Netherlands, Norway, Serbia and the United Kingdom — at least. The press room descends into a hubbub as reporters grab their notepads or leap to their laptops. The press officers trying to run the show look on, bemused.

13:50 Just before the closing ceremony starts, a television crew searches for a miserable American. Pluto, after all, was discovered at the Lowell Observatory in Flagstaff, Arizona, by the American Clyde Tombaugh. The search so far seems to have been fruitless. But I do see someone waving a picture of Pluto the Disney dog somewhere near the front...

14:35 "You will need a pen or a pencil," says Bell Burnell, who is chairing the session. The audience duly rummages in its bags, in order to add inverted commas to the category 'dwarf planets' and clarify the situation over satellites.

A speaker from the floor suggests, to much laughter, dropping all the resolutions except footnote 1 to 5A: "The eight classical planets are: Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus and Neptune."

14:43 At last, the vote. Astronomers wave little yellow cards in the air to indicate their support for resolution 5A — that's the one that recognizes three categories of object: planets, 'dwarf planets' and small Solar-System bodies. A few people wave their cards to vote the resolution down, a few abstain.

A moment's hesitation from the chair. Then: "I believe the resolution is clearly carried."

Amazing! A decision! I wouldn't have predicted that at the week's beginning.

Bell Burnell brings out teaching aids from under the table. A blue balloon to represent the planets. A stuffed Disney Pluto and a box of cereal (Ceres, therefore cereal, get it?) stand in for the 'dwarf planets'. There's something indistinguishable and lumpy for the small Solar-System bodies.

Next, a vote on resolution 5B. Are classical planets and 'dwarf planets' all planets proper, giving us two classes of planets and making 'planet' an umbrella term? (Out comes an umbrella labelled 'planets'.) Ninety-one in favour. The number against is overwhelming — no need to count again.

"It's clear that resolution 5B is not passed," the chair reports. So, we have eight planets only. Pluto is out.

Straight after the vote, I see Richard Binzel of the Massachusetts Institute of Technology, a member of the Planet Definition Committee. He says, with some relief, "it's over, it's done."

Oh no it's not. ■

**MOON SPECIAL**

Check the website from 1 September for slideshows, features and more.

www.nature.com/specials

NASA

Lunar probe ready to bite the dust

After nearly three years in space, Europe's Moon mission is out of fuel. On 3 September the spacecraft SMART-1 will smash into the lunar surface near the Lake of Excellence in the Moon's mid-southern latitudes. Astronomers hope the dust kicked up from the rocky site will provide information about the Moon's composition and impact history.

Researchers at the European Space Agency (ESA) are very happy with what the mission has achieved. The craft was the agency's test-bed for a new type of thruster called an ion engine. SMART-1 is the second such engine to be used in space — the first was on board NASA's comet-chaser Deep Space 1.

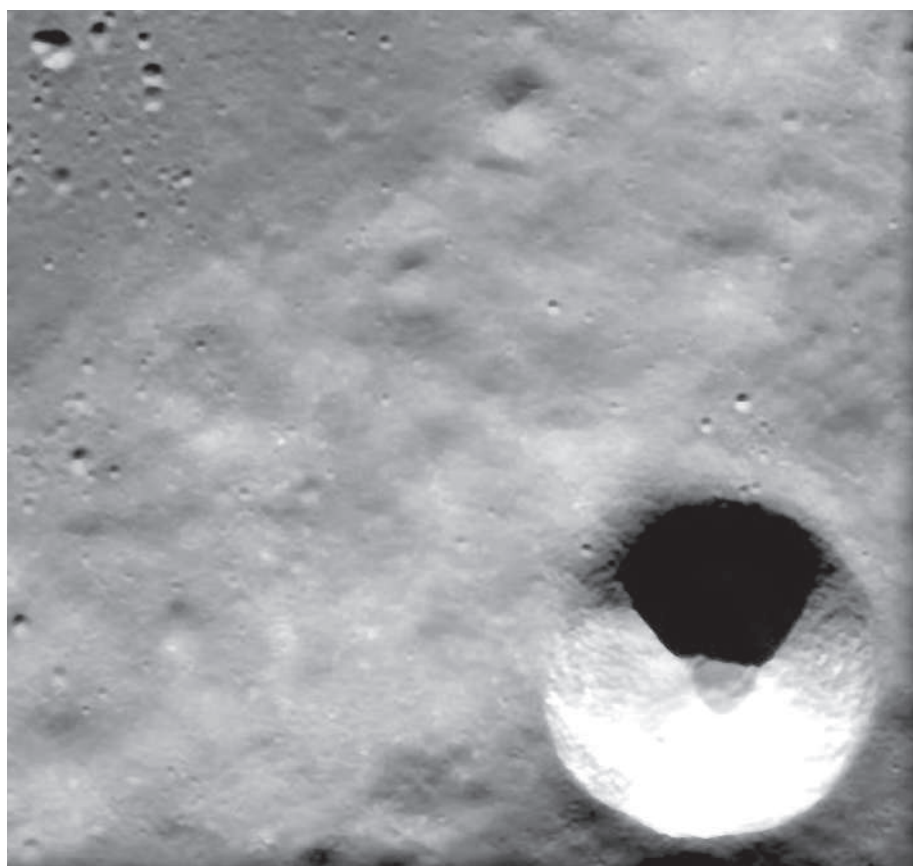
Ion engines work by accelerating a stream of ionized atoms through an electric field. The one on SMART-1 provides about as much force as a postcard resting in the palm of your hand. That made its journey painfully slow: it took 14 months and a roundabout trip of 84 million kilometres for it to spiral out of Earth's orbit and reach the Moon. In contrast, Apollo 11's conventional propellants allowed it to travel a roughly 400,000-kilometre line between Earth and the Moon in just over 3 days.

The upside is that SMART-1's journey used around 70 kilograms of fuel, about ten times less than would have been needed with regular propellants. So, although ion engines pick up speed quite slowly, they are extremely effective — and fast — for long-distance trips using light craft.

Learning to drive an ion engine is different from using chemical propellants, according to SMART-1 project scientist Bernard Foing. Chemical propellants are used in short burns lasting seconds or minutes. For those critical periods a craft must be precisely controlled, but otherwise it coasts passively to its destination. But ion drives are always on, so the course of the craft must be constantly monitored and corrected. "It's much more complicated than driving with chemical energy," says Foing.

Despite a few episodes in which the engine spontaneously switched itself off, Foing says that SMART-1 performed well — good news for ESA's upcoming Bepi-Colombo mission to Mercury in 2013. Bepi-Colombo will use conventional propellants to break out of Earth's orbit quickly, then an ion drive to speed towards the Sun. The engine's tiny but continuous thrust over the course of months should cut flight time in half.

Once SMART-1 arrived at its destination, it used a digital camera and two spectrometers to



ESA/SPACE-X

Moonshot: Europe's SMART-1 probe has offered astronomers a fresh view of the lunar surface.

survey the Moon's surface. The craft orbited for 22 months gathering data on the abundances of calcium, aluminium, magnesium and silicon on the lunar surface.

Several teams have been awaiting the release of the full results. For example, Wim van Westrenen, a petrologist at the Free University in Amsterdam, hopes they will help him investigate the Moon's origin. Most lunar researchers believe that the satellite formed when another body crashed into Earth billions of years ago. But it's not clear, for example, which of the two colliding bodies contributed most of the Moon's mass.

"There are dozens of models for where and how and what materials melted in the Moon," says van Westrenen. In a five-year project starting in January 2007, he hopes to work out how the process occurred by crushing various combinations of materials at the high temperatures and pressures that would have been present when the Moon formed, to try to recreate the

exact composition detected by SMART-1.

ESA's mission will be followed by a glut of lunar visits. In 2007, Japan is scheduled to launch SELENE, a 3-tonne spacecraft that will survey the Moon's mineralogy, topology and gravity gradients. The same year, India will launch Chandrayaan-1 and China will launch Chang'e 1 — both will be those countries' first missions beyond Earth orbit. In 2008 NASA will launch its Lunar Reconnaissance Orbiter to survey the Moon's surface.

The overlap probably has more to do with national pride than science. But Paul Spudis, a lunar specialist at Johns Hopkins University's Applied Physics Laboratory in Laurel, Maryland, says that the data will be welcome nonetheless. In the rush to beat the Soviet Union to the Moon, the Apollo programme did only piecemeal mapping of the satellite's surface. And after astronauts finished their visits, he says it was ignored for decades by planetary researchers. "Mars has better global imaging than the Moon does."

Geoff Brumfiel

"Using an ion engine is much more complicated than driving with chemical energy."

Plan B contraceptive gets green light from FDA

By approving the long-delayed Plan B emergency contraceptive, the US Food and Drug Administration (FDA) has restarted the stalled nomination of its likely next chief.

The 24 August decision means that women at least 18 years old will be able to buy the contraceptive over the counter. Approval of the drug had been held up for years because of opposition from conservative groups. In 2005, women's health advocates criticized the FDA's chief at the time, Lester Crawford, for overruling his scientific advisers and further stalling the decision on whether to make Plan B available without a prescription.

Andrew von Eschenbach, the cancer surgeon who is acting commissioner of the FDA, may now have a chance to take the reins permanently. His nomination had been held up by two senators angry over the Plan B delay, who now say they lift their objections.

Pro-evolution Vatican Observatory head replaced

A change of leadership at the Vatican Observatory has prompted speculation

among evolutionary scientists that the former head was removed because of his backing for evolution.

On 19 August, Pope Benedict XVI announced that George Coyne, who has headed the observatory for almost 30 years, would be replaced by José Funes, an extragalactic astronomer who has been associated with the observatory since 2000. Coyne caused controversy last summer

when he criticized an attack on evolution by Cardinal Christoph Schönborn of Vienna, a senior member of the Catholic Church.

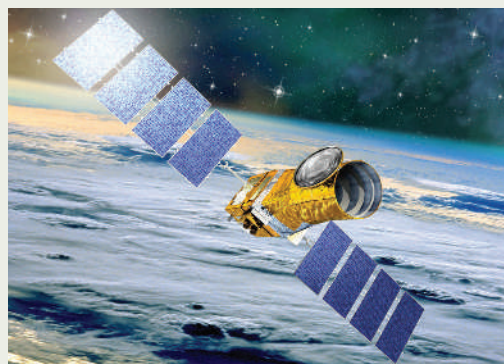
The Vatican observatory, which dates from 1891, has its headquarters at the papal summer residence in Castel Gandolfo, outside Rome. Research is carried out at the Vatican Observatory Research Group, which is based at the University of Arizona in Tucson.

Stargazers see trouble with revamped rocket

The launch of a planet-hunting, star-observing space telescope is in jeopardy because the rocket it was scheduled to fly on has yet to prove itself.

The COROT telescope will monitor changes in the brightness of distant stars to detect starquakes or planets passing in front of the stars. It was due to be hoisted into Earth orbit on 15 October by a Soyuz 2-1b launch vehicle, a revamped version of the successful Russian Soyuz boosters. But the contract arranged by the mission's consortium, led by the French space agency CNES, required the new launch vehicle to have made two successful flights before carrying COROT.

It now looks unlikely that this can happen in time, leading mission scientists to call for COROT to be put instead on a tried-and-trusted Soyuz Fregat as soon as the telescope can be reconfigured, probably in November.



D. DUCROS/CNES

Europe clamps down on rice imports from US

The European Union last week ordered that rice imports from the United States be certified as free from transgenic contamination. The move follows the discovery that traces of a herbicide-tolerant genetically modified strain, called LL Rice 601, were found in commercial samples of long-grain rice.

The ruling, expected to be in place for at least six months, aims to prevent the strain from entering rice stocks in Europe, which imports rice worth \$90 million from the United States each year. Japan has banned imports in response to the discovery.

The strain, made by German company Bayer CropScience, is not approved for sale in the United States, although two other engineered strains with the same herbicide-resistance protein are commercially available.

Katrina caused rise in mental-health problems

Cases of serious mental illness have doubled among survivors of Hurricane Katrina, which struck the US Gulf Coast a year



Trail of destruction: the effects of Hurricane Katrina run deeper than damage to property.

ago. But survivors do not think of suicide any more often than they did before the hurricane struck.

Researchers led by Ronald Kessler of Harvard Medical School studied 1,043 people in Louisiana, Mississippi and Alabama after Katrina hit. They compared that sample with 826 people in the same area interviewed between 2001 and 2003 for a national mental-health assessment. In the earlier study, 6.1% displayed serious mental illness, compared with 11.3% after the hurricane.

But thoughts of suicide were less frequent after the hurricane — perhaps reflecting

optimism that things could only improve, the researchers suggest. The findings were published online on 28 August in the *Bulletin of the World Health Organization*.

NASA counts down to next Moon craft construction

NASA is firming up its plans to build the next lunar spaceship. Dubbed Orion, it will be about five metres in diameter, with the same conical shape as the old Apollo spacecraft, but more than twice as spacious inside.

NASA plans to announce on 31 August who will build the spacecraft: either Lockheed Martin or a consortium consisting of Northrop Grumman and Boeing. The craft's new name, which was supposed to be announced on the same day, has already been accidentally revealed by astronaut Jeff Williams, currently aboard the International Space Station.

Orion will take over cargo and crew flights to the station no later than four years after the planned retirement of the space shuttle in 2010. Together with the Ares launch vehicles and a lunar lander, it is intended to carry up to four astronauts to the Moon by 2020.

BUSINESS

Mothers of invention?

Women academics are less likely than men to take out patents.

Emma Marris investigates the reasons why.

While running an analysis of academics' links with industry, two economists noticed something odd in their database of commercially well-connected scientists. Women were seriously underrepresented — even taking their minority status in academia into account.

"We were on the 208th line, and there was the first female name," says Toby Stuart, an economist at Harvard University. To put numbers on this apparent imbalance, Stuart and his colleague Waverly Ding of the business school of the University of California, Berkeley, started a new study, together with Fiona Murray of MIT, using patenting as a proxy for involvement with the commercial sector.

Their results appeared earlier this month in *Science*¹. Taking a 30-year period, the team looked at the patenting activity of more than 4,000 academics in patent-heavy life sciences such as molecular biology and microbiology. After removing confounding factors — such as publication productivity and differences between institutions — they found that, overall, women patented at around 40% the rate of the men. The discrepancy between men and women was, however, getting less over the years.

A similar gender gap was highlighted in two studies in the *Journal of Technology Transfer* in 2005, although technology-transfer professionals interviewed for this article appear not to have noticed. John Fraser, president of the Association of University Technology Managers, admits: "I guess I've been gender blind."

Tech-transfer specialists may have missed the gap because it gets lost in the gender imbalance in academia, particularly among senior faculty. But they are also more focused on the invention. "If the ideas are good, no matter who they come from, business is going to say, 'hey we want it', and do a deal," says Fraser.

Stuart sums up the reasons for the discrepancy in two words: attitudes and networks. "We go all the way back to the 1970s, in the recombinant DNA period, when patenting was really frowned upon," he says. In interviews, Stuart found that older women expressed a lot of reservations about patenting. They knew their careers weren't as secure, so were hesitant about doing something that ran counter to the norms of the



To patent or not? Many women scientists find themselves too busy.

group. "And the men just seemed to be much better networked into industry," he says.

Laurel Smith-Doerr, a sociologist at Boston University and co-author of one of the *Journal of Technology Transfer* studies, has another theory². She looked at patent quality, as well as quantity, in the life sciences: "We also find that women patent less, but for women who patent, their patents are cited more often and more widely — in more areas of the life sciences. So perhaps they are patenting for quality."

The other *Journal of Technology Transfer* article sheds light on where women fall out of the patenting pipeline³. Jerry Thursby, an economist at Emory University in Atlanta, and his wife Marie Thursby, an economist at nearby Georgia Tech, studied the gender difference in disclosures — the paperwork that is filed

with the university on a possibly patentable invention. "We were trying to figure out who is interested in licensing," says Jerry Thursby. Their result? A man is 43% more likely to have disclosed an invention to his tech-transfer office. This means that many women aren't even beginning the process.

But perhaps the most compelling observation comes from Kjersten Bunker Whittington, Smith-Doerr's co-author, who is researching the question for her dissertation. She finds that women without children patent at the same rate as men. So the gender gap is really a gap between women with children, and everyone else.

Jennifer West, a bioengineer at Rice University in Houston, Texas, has more than a dozen patents, and was surprised at the difference detected in these studies. "This is an area of my career that I hadn't thought gender came into. It was instilled in me even as an undergraduate that it was my responsibility to patent."

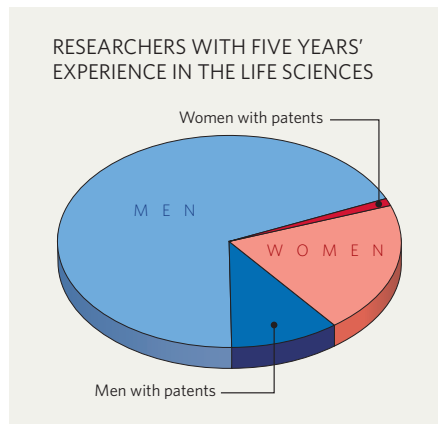
West admits, though, that it takes time, and is not currently as highly

valued in tenure and promotion decisions as a publication record. For women balancing children and a career, it may therefore look like an optional extra. "If junior faculty are trying to balance their family lives, patenting isn't necessarily the best use of their time," she says.

The implications of a patenting gap may be bad for women if patents become part of how academics are evaluated. But not everyone is convinced that will happen. At the moment, only a small fraction of academics ever patent. "There seems to be this growing myth that everyone is constantly running to the technology office," says Thursby. In practice, he says, a very small number of people are responsible for most of the patents.

The real price being paid may be the comparative exclusion of women scientists from business opportunities. "Patents really are a precursor to involvement at multiple levels in a company," says Stuart, "and we need more women in high levels in high-tech companies."

If childcare really is the limiting factor, it adds to the list of arguments for better, cheaper daycare for academic parents. Smith-Doerr thinks the benefits to society are worth the price. "I believe the more diverse the group of people who are developing knowledge, the better that knowledge will be," she says. ■



1. Ding, W. et al. *Science* **313**, 665–667 (2006).

2. Bunker Whittington, K. & Smith-Doerr, L. *J. Technol. Transfer* **30**, 355–370 (2005).

3. Thursby, J. & Thursby, M. *J. Technol. Transfer* **30**, 343–353 (2005).



Fertility on a shoestring

IVF isn't something most Westerners associate with Africa. But low-cost methods are urgently needed to treat the misery of infertility rampant on the continent, says **Helen Pilcher**.

Several years ago, Betty Chishava was thrown out of her family home in Harare, Zimbabwe, because she failed to fall pregnant and didn't want to sleep with her husband's brother. Desperate for an heir and a cure for the stigma of infertility, her husband Herbert took a new wife. Betty was left penniless and alone.

Betty's story is played out in millions of homes across sub-Saharan Africa, where up to one-third of couples are infertile¹ and the pressure to produce children is immense. "In Africa, a woman's worth is defined by her fertility," says Chishava. As the years roll by and a couple's lack of children becomes all too apparent, personal tragedies turn into public humiliation and shame. Perceived as evil or cursed, a woman without a child may be beaten and is commonly ostracized by family and friends. Some risk the threat of HIV to conceive through sex with multiple partners. Some fake pregnancies and steal newborn babies. Some just can't bear it any more and take their own lives.

"If you're a woman in sub-Saharan Africa and you don't have a child, you're worth less than a dog," says fertility specialist Willem Ombelet from the Genk Institute for Fertility Technology in Belgium, who has worked in

Africa for more than two decades.

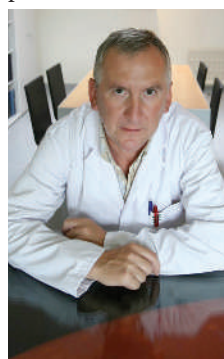
Treatments for infertility are becoming available in the developing world — but slowly. In 1989, little more than a decade after the world's first baby conceived through *in vitro* fertilization (IVF) was born in Britain, western Africa hailed its first IVF success: a boy born at Lagos University Teaching Hospital in Nigeria. Since then a sprinkling of private fertility clinics has sprung up, offering high-tech treatments from the developed world. But the therapy is too expensive for most Africans.

Initiatives are afoot to make fertility treatment more accessible. A number of scientists have proposed methods for developing simpler, low-cost alternatives to the high-tech

drugs and equipment currently used for fertility treatments. And others are working to prevent the sexually transmitted infections that account for most cases of infertility. But even as campaigners strive to make these approaches a reality, they face daunting prejudice — in both Africa and the wider world.

Perhaps the biggest stumbling block is the insidious conviction (in Western circles) that sub-Saharan Africa simply cannot have an infertility problem. "Governments worldwide put money into family planning in the developing world, but no one wants to focus on infertility," says Ombelet. The average couple in that region has five or six children², so many governments think that being too fertile should be the focus. And they find it hard to justify expensive fertility treatments in settings with few resources that have more obvious problems, such as malaria and HIV. But scientists such as Ombelet insist that the problem needs to be addressed. "Infertility is consistently overlooked in aid and development work," he says. "We have to convince the Western world that infertility in Africa is a real problem."

The sad thing is that much of Africa's infertility could be easily prevented, as infections are the main cause of infertility on the continent. Infections such as gonorrhoea and chlamydia



"We have to convince the Western world that infertility in Africa is a real problem."

— Willem Ombelet

A. ROSS/GETTY IMAGES



Bundles of joy: Kenya celebrated its first IVF babies last year and clinics are springing up across Africa.

often go untreated, spreading to the reproductive tubes where they cause blockages and scarring. The acceptance of one simple thing — the condom — could change it all.

But cost and cultural taboos restrict condom use. Women, most of whom depend upon men for economic security, find it hard to negotiate safe sex and difficult to refuse intercourse; those who do may risk a beating. Less than 2% of married women in Africa use condoms³.

Bodies such as the World Health Organization have programmes aimed at improving gender equality and the availability of contraception. And results are shortly expected from clinical trials of microbicide gels, which may be effective against a range of sexually transmitted diseases⁴.

Botch jobs

Meanwhile, the unplanned pregnancies that result from a lack of contraception pose another threat to a woman's fertility. General contraceptive uptake among women remains low, averaging 23% across sub-Saharan Africa as a whole, reaching just over 50% in South Africa, and dropping to less than 10% in Nigeria³. All too often, deliveries and abortions turn into botch jobs that cause infection, says obstetrician and gynaecologist Osato Giwa-Osagie from Lagos University Teaching Hospital. He says this is the second biggest factor in the region's female infertility.

Female genital mutilation, which is more common in Africa than anywhere else, carries similar risks. Up to 140 million women have had part or all of their external genitalia removed⁵. The practice is usually performed by traditional doctors with crude instruments and without anaesthetic. Resulting infections can spread internally.

Men, of course, can also be infertile. "But

it's always the women who are blamed," says Chishava. Male infertility accounts for up to 40% of cases of childless couples⁶. Men are usually born with their fertility problems, although infections are also a factor. And yet male infertility is so taboo that no one will admit it exists. Families go to great lengths to cover it up: some resort to the traditional practice of getting a husband's brother to impregnate his wife, something Chishava refused to agree to. Most men in Zimbabwe, and some other countries including Nigeria, would rather change their wife than admit to an infertile marriage, she says.

Cheap tactics

Changing age-old prejudice is going to take a while. In the meantime, scientists and doctors in Africa and elsewhere are turning to IVF and other fertility treatments to help.

Africa has a more-than-respectable history in assisted reproductive technologies (ARTs). After Nigeria's IVF success in 1989, Giwa-Osagie expected African governments to increase public spending on ARTs. "In terms of technology, we were just a few years behind Britain," he says. But the people with the purse strings prioritized other health concerns, such as malaria and diarrhoea. Africa's public infertility clinics began to feel the pinch and close



"In terms of technology, we were just a few years behind Britain in 1989."
— Osato Giwa-Osagie

down. "The service became very fragmented and many couples ended up going from doctor to doctor," says Nigerian fertility specialist Richard Ajayi. Those who could afford it, travelled abroad for treatment.

Ajayi and others like him turned to the private sector for funding. Venture capital secured, in 1999 Ajayi founded the Bridge Clinic in Lagos. With its recently founded sister clinic in Port Harcourt, Ajayi's clinics perform more than 500 cycles of IVF treatment a year.

The Bridge Clinic is a polished outfit with state-of-the-art laboratories and new equipment; other places have more humble beginnings. In 2003, gynaecologist Edward Sali set up the Kampala Gynaecology, Fertility and Maternity Centre in Uganda. Funds were tight, so Sali turned his bedroom into the laboratory. "We had to move out," he says.

Clinical excellence

There are now more than two dozen private fertility clinics scattered across nine or more sub-Saharan African countries. Virtually all forms of ART practised in the West are available, with IVF and artificial insemination by a husband's sperm the most common. Thanks in part to collaboration with sister clinics in the developed world, success rates are approaching those seen in the West. But there's a problem — the cost.

In Nigeria and other countries in sub-Saharan Africa, a single IVF treatment costs around US\$2,500. But the minimum wage in Nigeria is just US\$52–60 a month⁶ and there are a great many people scraping by on a dollar a day, or less. This makes IVF and other techniques too pricy for most. Ajayi estimates that only 5–10% of those who could benefit from fertility therapy can afford private treatment.

So what is to be done? Fertility expert Alan Trounson from Monash University in Melbourne, Australia, thinks it's time to go back to basics. He believes that the cost of IVF could be slashed by replacing expensive drugs and high-tech equipment with safe, low-cost alternatives. And he's drawing on experience from veterinary medicine, paediatrics and the early days of IVF to do just that.

Normally when a woman undergoes IVF, she is first injected with hormones called gonadotropins, to help her produce more eggs. Shortly before the eggs are harvested, the woman receives another hormone injection to help the eggs mature. The eggs are then collected using ultrasound guidance, and fertilized in the lab. The work is carried out inside a sterile cabinet called a laminar flow hood, before being transferred to a humidified, gas-filled incubator where the fertilized eggs are left to divide for a few days before being implanted.

"Over the years, IVF has become tailored to treat the Harley Street end of the market," says Trounson. "But there's no reason it can't be tweaked and simplified to create something that is affordable and safe, with a reasonable output."

BRIDGE CLINIC



The Bridge Clinic in Lagos uses high-tech tools, but a converted humidicrib (right) also provides a sterile environment in which to manipulate embryos.

Costly hormone injections could be replaced with cheaper alternatives. Most women use 30 ampoules of gonadotropin per treatment cycle, resulting in about a dozen eggs: the total cost is US\$300–450. But the dose could be reduced or replaced with clomiphene citrate. This drug, which was routinely used in IVF treatments in the late 1960s, also stimulates ovulation but produces fewer eggs. And 15 tablets cost about US\$1.

Softly, softly

Critics of clomiphene citrate caution that it can sometimes damage the uterus lining, making embryo implantation less likely. But advocates say that the 'softer' drug is less likely to trigger ovarian hyperstimulation syndrome, a rare side effect of gonadotropin therapy that can cause diarrhoea, vomiting and breathing difficulties. In the early 1980s, Trounson managed a roughly 5% live birth rate using clomiphene citrate, and antenatal care and tissue-culture techniques have come a long way since then. Research is ongoing, but Ombelet estimates that three cycles with clomiphene citrate should be as effective as two cycles with gonadotropins.

Slashing the price of drugs means nothing, however, if doctors cannot afford the equipment needed to fertilize and nurture the eggs. Western IVF laboratories are replete with technology that costs tens of thousands of dollars, but much of it can be done away with.

In place of the laminar flow hood, Trounson suggests using a 'humidicrib' — a plastic box

more commonly used for keeping newborns snug. It's a tenth of the price and can be modified to create a portable, near sterile environment in which to handle embryos.

And instead of incubating the embryos with carbon dioxide from an expensive cylinder, Trounson recommends exhaling across the culture media before sealing it in a plastic bag, a technique commonly used in veterinary IVF. Then remove the need for an incubator by dropping the bag containing the the Petri dish into a warm water bath. Such 'submarine incubators' have been used for cow embryos for more than a decade. "People didn't think to use it in an IVF setting because it's not seen to be sophisticated enough," says Trounson.

Pilot studies are needed to assess the safety and efficacy of such low-cost IVF protocols. Unfortunately, a lack of awareness, cash and political push means this is not happening. So



A. TROUNSON

Trounson and others are lobbying hard to put African infertility on the international agenda. Ombelet is organizing a meeting in Arusha, Tanzania, in February 2007, where scientists, clinicians, ethicists, policy-makers and women's organizations will draw up a plan of action. And the Nigerian Fertility Society is drafting guidelines on infertility treatments, which it hopes will be accepted by its government.

As awareness increases, it is hoped that government money will trickle back to fund public infertility clinics. Such low-cost IVF could help millions of women. But government support and publicly funded clinics will mean little if the social stigma is not tackled. "Women have no voice in Africa," says Ombelet, "so raising the status of women may be the hardest job of all."

Home truths

That's where Chishava comes in. When she was thrown out of her family home, she realized that there was a need for female counselling and education. So six years ago, she set up the Chipo Chedu Society (meaning 'our gift' in her language). The organization aims to help childless women become financially

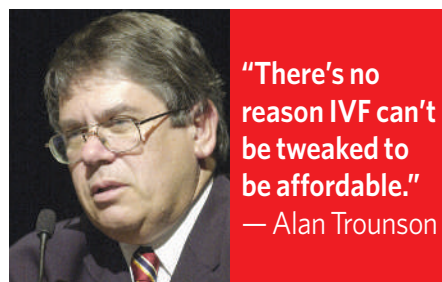
independent, teaching them practical and business skills such as batik and bookkeeping. It puts women in touch with medical experts and fights prejudice through rural workshops on infertility. The society has more than 500 members and Chishava hopes to see her network spread across Africa.

Chishava has since been reunited with her husband. His second and third marriages failed to produce children, and he gradually accepted that he was infertile. He apologized to Chishava and fully supports her work. Now 54, Betty is mother to five children — all given to her by family members — but she would dearly love to have a child of her own and is intrigued by IVF. Until that happens, Chipo Chedu is her true baby. "I will only find peace of mind when the programme flourishes," says Chishava. "If my neighbour has no children, then their problem is my problem."

Helen Pilcher is a science writer based in Nottinghamshire, UK.

1. Daar, A. S. & Merali, Z. Infertility and Social Suffering: the Case of ART in Developing Countries (World Health Organization, 2001); available at <http://www.who.int/reproductive-health/infertility/5.pdf>
2. UNICEF Fertility and Contraceptive Use (2001); available at <http://www.childinfo.org/eddb/fertility/index.htm>
3. UN World Contraceptive Use (2003); available at <http://www.un.org/esa/population/publications/contraceptive2003/wcu2003.htm>
4. Pilcher, H. *Nature* **430**, 138–140 (2004).
5. World Health Organization factsheet 241 (2000); available at <http://www.who.int/mediacentre/factsheets/fs241/en>
6. Giwa-Osagie, O. F. ART in Developing Countries with Particular Reference to Sub-Saharan Africa (World Health Organization, 2001); available at <http://www.who.int/reproductive-health/infertility/6.pdf>

See also Editorial, page 957.



"There's no reason IVF can't be tweaked to be affordable."
— Alan Trounson

AAP IMAGE/JULIAN SMITH

SICK SEAS

The rising level of carbon dioxide in the atmosphere is making the world's oceans more acidic. **Jacqueline Ruttimann** reports on the potentially catastrophic effect this could have on marine creatures.

It's not hard to imagine a tonne of water: it is a week's worth of not-very-deep baths. Getting to grips with a billion tonnes of water is more of a challenge. That would be a similar bath for every man, woman and child on the planet; a week's worth of flow for the Nile. To really expand your mind, go further still, to a billion billion tonnes — enough water to give every human a day's worth of the Nile instead of a shallow bath. There are dwarf planets that weigh less than a billion billion tonnes. Yet Earth's oceans weigh more.

If it is hard to imagine something so vast, it is perhaps even harder to imagine changing it. But humanity is changing the oceans. From the tropics to the Arctic, the seas are sucking up human-driven emissions of carbon dioxide — about half of the excess belched into the atmosphere over the past two centuries from fossil-fuel burning and cement manufacturing plants¹. When carbon dioxide dissolves in water, carbonic acid is produced: as a result the oceans are becoming more acidic. "It's basic chemistry," says Joanie Kleypas, a marine ecologist at the National Center for Atmospheric Research in Boulder, Colorado. "It's hard to say that this is not happening."

Over the past few years, scientists have documented how increasingly acidic seas could eat away the armour of many creatures — blunting the spikes on sea urchins and dissolving the covering on corals. In artificially acidified waters, some animals, such as squid, have problems swimming because the corrosive water affects their respiration rate. Others, particularly tiny organisms with carbonate shells, lose their protective shields as the acid eats away at them². But research into low-pH oceanography is, as yet, sparse. "We're just starting to grapple with what low pH will mean for ocean communities," says Jim Barry, a marine biologist at the Monterey Bay Aquarium Research Institute (MBARI) in Moss Landing, California.

This ocean acidification is unlike the atmospheric warming also being caused by carbon dioxide in that it is fairly predictable; plotting its future course requires little more than school chemistry, as opposed to sophisticated modelling. The rate of acidification is pretty much

unprecedented. Before the industrial revolution, the rise in the level of carbon dioxide in the atmosphere was relatively slow — giving oceans time to circulate the waters being made more acidic in the shallows with acid-neutralizing carbonate sediments in the depths.

In the past few decades, carbon dioxide has been building up far more quickly, and the ocean is becoming acidified at a rate that outpaces the action of sedimentary antacids. The rate of change is perhaps 100 times anything seen in the past hundreds of millennia, as suggested by isotope studies of ancient sediments. In the century to come, sea creatures will find themselves in conditions that their ancestors never had to face. These organisms have never been forced to adapt to lower pH, says Ulf Riebesell, a marine biogeochemist at the Leibniz Institute for Marine Sciences in Germany. "They've never seen this before in their evolution."

Acid attack

The acidified waters eat away at the carbonate skeletons that protect many marine organisms. By some estimates, calcification rates will decrease by up to 60% by the end of this century. If so, carbon dioxide in the ocean could represent a chemical threat to the biosphere as severe as that posed by the build up of carbon dioxide in the atmosphere.

A 2005 report from the Royal Society in London called for more research to help quantify the threat. Another report, released last month by three US research agencies, the National Science Foundation, the National Oceanic and Atmospheric Administration, and the US Geological Survey, laid out research strategies for tackling the problem. And the Intergovernmental Panel on Climate Change (IPCC) — the international body tasked with quantifying the effects of climate change — is flagging ocean acidification as a problem for the first time in its next report, due in early 2007.

In 1800, the carbon dioxide in the atmosphere was 280 parts per million, and the oceans' pH averaged 8.16. Today there are 380 parts per million of carbon dioxide in the atmosphere, and the pH of the oceans is on average 8.05.



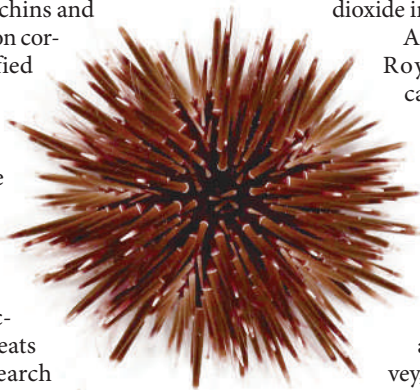
Estimates suggest the pH could drop to 7.9 by the end of the century. The drop from 8.16 to 7.9, says Kleypas, "doesn't sound like much, but it's a lot." Each one-step change on the pH scale indicates a tenfold change in acidity.

Just as terrestrial ecologists experiment by growing plants in high concentrations of carbon dioxide, so marine biologists are beginning to investigate what life in the less-alkaline ocean may be like.

In a fjord in southwest Norway, Riebesell has set up an outdoor laboratory consisting of a raft with what look like giant milk cartons moored to it. The containers, known as 'mesocosms', are 50-litre vessels filled with coccolithophores — photosynthesizing plankton, or phytoplankton, with carbonate coverings. Riebesell immerses the coccolithophores into tanks that are aerated with the projected levels of carbon dioxide in the next 50 and 100 years. He calls them "the oceans of the future".

The coccolithophores' outer casings — tiny hubcaps known as coccoliths — are made of the carbonate mineral calcite. Riebesell found that exposing coccolithophores to three times the present-day atmospheric level of carbon dioxide caused nearly half of their protective coating to disintegrate³.

Such changes don't bode well for one of the ocean's most abundant types of phytoplankton,





M. USHODA/IMAGE QUEST MARINE

Fragile lives: acidified seas could dissolve corals and blunt the spines of sea urchins (left).

and what's bad for phytoplankton is likely to be bad for the food webs they sustain. By aggregating on the surface of waste from the upper levels of the ocean (mainly fish droppings), the coccoliths help it to sink down to the seabed communities, which recycle its nutrients into the ocean. Weakening the coccoliths could have knock-on effects on nutrient cycling.

Most at risk

Riebesell is now looking at the effects of increasing acidity on plankton eggs and larvae; other researchers are studying the dangers posed to larger animals, such as sea urchins and snails. In one experiment, marine biologist Yoshihisa Shirayama, at Kyoto University in Japan, placed sea urchins and snails separately into 30-litre tanks filled with water containing 550 parts per million of carbon dioxide. Within three months, the creatures dropped in weight by roughly 8%. Their calcite spines became blunted, says Shirayama, and so brittle that they broke off easily when handled.

If carbon dioxide levels in the tanks were quickly lowered, the animals eventually recovered. If they were exposed to high levels over the long term, such as for a year, most of them died⁴. "It's obvious that animals cannot adapt to large

change in carbon-dioxide concentration," says Shirayama.

Other disturbing news comes from experiments led by MBARI's Barry. For most of his career, carbon dioxide has just been a tool: he used to carry a fire extinguisher to intertidal zones so he could bubble the gas through the sea water, put the animals living in it to sleep and thus capture them more easily. Now he's looking at carbon dioxide as a threat.

Working with his post-doctoral fellow Eric Pace, Barry has been studying how two crab species respond to elevated levels of carbon dioxide. Preliminary results suggest that the shallow-water Dungeness crab, (*Cancer* species), does much better than the deep-sea Tanner crab (*Chionoecetes* species), as measured by how long it takes their metabolism to recover from a high dose of carbon dioxide. That difference points out a key point about the acidifying oceans, says Barry: "Deep-sea animals are probably much more sensitive to changes than animals found in shallow water."

Barry and Pace presented their findings on

the crab studies in July, at a deep-sea biology symposium in Southampton, UK. In related work, Barry has found that microscopic animals living in seafloor sediments seem to decline after being exposed to a drop in pH of just 0.1 for one month. Oddly, they seem to recover after a week or two — perhaps, Barry says, because fewer of them are being eaten by larger animals, which may have died off, or because they have more to eat themselves as the dead bodies of other small creatures, killed by the pH change, drift to the sea floor.

Mass die-off

Also at risk in the deep sea are cold-water corals, some of the slowest growing corals on Earth. Corals are particularly vulnerable to acidification because reefs are built out of aragonite, a carbonate mineral that is more soluble than the calcite used by coccolithophores and sea urchins. "Deep-water corals could soon be in trouble," says James Orr, an oceanographer at the Laboratory for Climate and Environmental Science in Gif-sur-Yvette, France. By the end of the century, two-thirds of deep-water corals — as opposed to virtually none today — could be exposed to sea water that is corrosive to aragonite⁵.

Because carbonate is more soluble in high pressures and cold water, increased acidity has a greater effect at depth. But Orr's studies suggest that aragonite will be unstable at all depths throughout the Southern Ocean by 2100. Once the corals go, so too may other species that rely on them for resources such as shelter.

Despite such dire models, not all scientists are convinced that ocean acidification will be a major problem. In May, a paper in *Geophysical Research Letters* suggested that the rising carbon dioxide would have minimal biological impact in the ocean⁶. Hugo Loaiciga, a hydrologist at the University of California, Santa Barbara, argued that sediment carbonates would be able to buffer the predicted acidification.

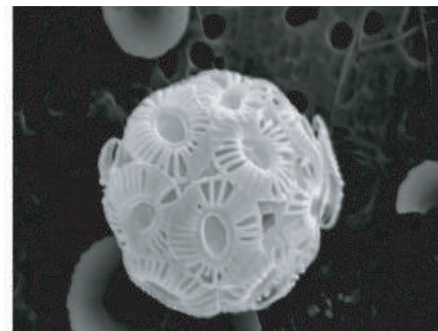
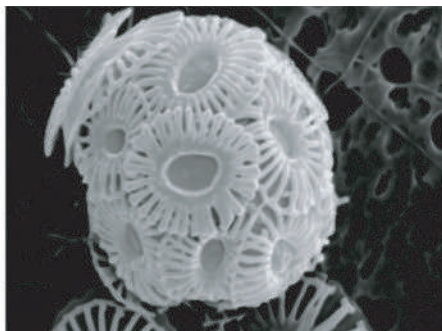
But it seems most experts disagree. A response, signed by 25 leaders in the field of ocean acidification, rebutted each of Loaiciga's points in detail. The experts also argued that many marine organisms would be sensitive to a drop of 0.2 pH units — a change Loaiciga had said would be essentially unimportant.

Ken Caldeira, an Earth-systems modeller at the Carnegie Institution of Washington in Stanford, California, who led the response, says the public needs to know now that this is a problem, and that it is unprecedented in its scale. In one model, Caldeira and his colleague Michael Wickett from the Lawrence Livermore National Laboratory, California, simulated the effects of an unregulated pulse of fossil-fuel burning that peaks around the year 2100.

The atmospheric carbon-dioxide level peaks shortly after the highest rate of burning. Then

"People should know that the consequences of what we're doing in the next decade will last for thousands of years." — Ken Caldeira

J. CUBILLOS



Lost protection: making sea water more acidic (centre and right) dissolves the outer casings of coccolithophores, tiny plankton that form the basis of food webs.

it slowly subsides as the carbon is picked up by the ocean — which slowly becomes more acidic at depth. At the surface, the pH might stop falling by 2750. A kilometre or so deep, however, the pH would still be dropping at the beginning of the next millennium⁷. “People should know that the consequences of what we’re doing in next decade will last for thousands of years,” says Caldeira.

In the meantime, researchers have a great deal to do trying to understand what the first decades of change could bring about. “So few studies have been done,” says Victoria Fabry, an oceanographer at California State University in San Marcos. “We’ve really just scratched the surface.” As an example, Fabry says lab studies on calcifying plankton are in such short supply that researchers have investigated only about 2% of all species.

The July report produced by the US research

agencies suggests paths for future studies. These include merging experiments done in the laboratory with those in the field, as well as identifying the effects of acidification on key-stone species such as coral and plankton and how changes to such species could affect the rest of the ecosystem. The report also notes the need to study the problem at all spatial scales; currently, most studies are focused on either the large-scale biogeochemical level or effects on individual organisms. Studies of ecosystems are in short supply.

A world of problems

To try to rectify that, some researchers are pushing for large-scale field experiments, such as marine versions of the open-air experiments in carbon-dioxide enrichment that have been taking place in forests for many years. Barry and his MBARI colleagues are working to

develop one of these experiments in the free ocean — it will be essentially a deep-sea cage set in a plume of water enriched in carbon dioxide.

In addition, disciplines are starting to merge. Oceanographers are beginning to collaborate with chemists to quantify the acidification reactions, and with engineers to develop robotic submersibles to collect deep-sea organisms for study. Social scientists are starting to get involved as well. The next Pacific Science Congress, to be held in Okinawa, Japan, in June 2007, will bring together natural scientists and social scientists to discuss how residents of coastal areas might adapt.

With more attention on the problem, a new possibility has raised its head. Ocean acidification might not just run in parallel with global warming — it could amplify it. The chalky coccolithophores, when blooming, lighten the surface of the oceans, which means more sunlight is reflected into space. Reduce their number and even if other phytoplankton take their place, that lightness will be gone.

Coccolithophores are also responsible for many of the clouds over oceans. They produce a lot of dimethylsulphide, which accounts for much of the aerosolized sulphate in the atmosphere above the oceans. Sulphate particles act as ‘seeds’ around which cloud droplets grow⁸. Remove them, and you could remove a significant fraction of the world’s clouds, warming the planet yet further⁹.

Human imagination cannot easily cope with the vastness of the oceans, or the complexities that change within them can bring about. But human industry faces no such obstacle in making change unavoidable for centuries to come.

Jacqueline Ruttimann is a freelance writer in Maryland.

Additional reporting is by Alexandra Witze.

U. RIEBESELL



Artificially acidified tanks of sea water are shedding light on the effects of lower pH on sea creatures.

1. Tedesco, T., Feely, R. A., Sabine, C. L. & Cosca, C. E.; available at www.research.noaa.gov/spotlite/archive/spot_gcc.html (2005).
2. Orr, J. C. *et al.* *Nature* **437**, 681–686 (2005).
3. Riebesell, U. *et al.* *Nature* **407**, 364–367 (2000).
4. Shirayama, Y. & Thornton, H. J. *Geophys. Res.* **110**, C09S08 (2005).
5. Guinotte, J. M. *et al.* *Front. Ecol. Environ.* **4**, 141–146 (2006).
6. Loàiciga, H. A. *Geophys. Res. Lett.* **33**, L10605 (2006).
7. Caldeira, K. & Wickett, M. E. *Nature* **425**, 365 (2003).
8. Charlson, R. J. *et al.* *Nature* **326**, 655–661 (1987).
9. Gunson, J. R. *et al.* *Geophys. Res. Lett.* doi:10.1029/2005GL024982 (2006).

A global initiative on sharing avian flu data

SIR — The global spread of the H5N1 avian influenza virus has already extensively damaged economies worldwide and food safety in developing countries. The spread of infection to new ecosystems results in adaptation of the virus to new hosts, including humans, which amplifies the potential for a flu pandemic. Because it is recognized that avian influenza viruses may be the progenitors of the next human pandemic virus, their genetic evolution should be tracked in detail and promptly investigated.

The full support of the international scientific community is therefore urgently needed to understand better the spread and evolution of the virus, and the determinants of its transmissibility and pathogenicity in humans. This in turn demands that scientists with different fields of expertise have full access to comprehensive genetic-sequence, clinical and epidemiological data from both animal and human virus isolates.

Several countries and international agencies have recently taken steps to improve sharing of influenza data^{1–4}, following the initiative of leading veterinary virologists in the field of avian influenza. The current level of collection and sharing of data is inadequate, however, given the magnitude of the threat. We propose to expand and complement existing efforts with the creation of a global consortium — the Global Initiative on Sharing Avian Influenza Data (GISAID; <http://gisaid.org>) — that would foster international sharing of avian influenza isolates and data.

Scientists participating in the GISAID consortium would agree to share their sequence data, to analyse the findings jointly and to publish the results collaboratively. Data would be deposited in the three publicly available databases participating in the International Sequence Database Collaboration (EMBL, DDBJ and GenBank) as soon as possible after analysis and validation, with a maximum delay of six months. The six-month deadline for data release is expected to become shorter as the consortium gains experience and works out its operating procedures.

GISAID's policies for rapid and complete data release are modelled on those established for community resource projects. These policies have successfully been employed previously, for example by the International HapMap Project (www.hapmap.org) — a project to map, and make freely available, data on DNA sequence variations in the human genome.

The GISAID consortium will comprise scientists from around the world working in the fields of animal and human virology, epidemiology and bioinformatics, as well as experts in intellectual-property issues. An

international panel of distinguished scientists will be formed to govern the charter and to advise the consortium.

As an international collaborative effort, GISAID offers many benefits to the world as a whole, as well as to individual scientists and to groups participating in the consortium. It would encourage valuable collaboration among researchers in industrialized countries and in the developing countries that are hit hardest by avian influenza. It would also attract international attention to the need for increased funding and technical assistance to help affected countries build comprehensive and sustained disease-surveillance programmes.

Peter Bogner*, **Ilaria Capua†**, **Nancy J. Cox‡**, **David J. Lipman§** and others||

*Chief executive, the Bogner Organization, 927 Fifteenth Street, Santa Monica, California 90403, USA

†Chair, Scientific Committee of the OFFLU OIE/FAO Network, Istituto Zooprofilattico Sperimentale delle Venezie, Viale dell'Università 10, 35020 Legnaro, Padova, Italy

‡Director, Influenza Division, Centers for Disease Control and Prevention, 1600 Clifton Road, Atlanta, Georgia 30333, USA

§Director, National Center for Biotechnology Information, National Institutes of Health, 8600 Rockville Pike, Bethesda, Maryland 20894, USA

||A full list of signatories to this letter as of 23 August 2006 is available as supplementary information at <http://www.nature.com/nature/journal/v442/n7106/extref/442981a-s1.pdf>.

1. *Nature* **441**, 1028 (2006).
2. OFFLU Keeps its Pace on Global Sharing Virus Samples Press Release (OIE/FAO Network of Expertise on Avian Influenza, 21 July 2006); available at <http://www.offlu.net/portals/0/pdf/Press.pdf>
3. Rukmantara, Tb. A. Bird Flu Data Now Open to All The Jakarta Post (4 August 2006); available at <http://www.thejakartapost.com/yesterdaydetail.asp?fileid=20060804.H07>
4. CDC and APHL Make Influenza Virus Sequence Data Publicly Accessible Press Release (Centers for Disease Control and Prevention, Atlanta, Georgia, 22 August 2006); available at <http://www.cdc.gov/od/oc/media/pressrel/r060822.htm>

Published online 24 August 2006.

Offsets could mitigate damage to biodiversity

SIR — The authors of the Commentary “Diversity without representation” (*Nature* **442**, 245–246; 2006) highlight the world's ineffectual response to the crisis of biodiversity loss. I agree that an international body similar to the Intergovernmental Panel on Climate Change would be useful. But I believe that what is most important is for individuals, corporations and governments to take responsibility for the impact of their activities on biodiversity, using the ‘avoid, mitigate and compensate’ hierarchy from the Convention on Biological Diversity's voluntary guidelines for impact assessment, to ensure that their actions do not result in net biodiversity loss.

The main obstacle to such action has been the lack of standards to measure impacts and to identify appropriate compensation. Recently, however, considerable progress has been made on biodiversity offsets, a process by which impacts are measured, then conservation investments made to ensure that the net impacts to biodiversity are neutral or even positive.

The most advanced analytical work involves wetlands in the United States, where any impact must be offset by restored wetlands that are equivalent in function to that lost. But leading companies are taking steps to offset their impacts in other ecosystems. For example, Peru's forests of *Polylepis* — home to several endangered bird species — have been reduced in size by 97% since the time of the Incas, especially in recent decades by grazing and fuelwood collecting. Now the company Minera Antamina has undertaken to protect and reforest an area hundreds of times larger than the area of *Polylepis* affected by its mine (see <http://biodiversityneutral.org>).

Although regulations requiring offsets are essential to halt the loss of habitat, existing frameworks are sufficient for individuals and corporations to take voluntary responsibility for the loss of biodiversity caused by their activities. And should an intergovernmental panel on biodiversity be established, offsets would provide an immediate focus for research and policy-making.

Art Blundell

Biodiversity Neutral Initiative, 122 Haida Trail, Nanaimo, British Columbia V9S 3G1, Canada

Funders should allow for cost of publication

SIR — Open access to the literature allows scientists in the developing world to read original research papers for free, which contributes to scientific advancement. Nonetheless, in these same countries, funds are not sufficient to pay the publishing charges made by some publications, including ‘open access’ journals. For this reason, many journals waive fees for scientists from developing countries who submit to them.

Although these waivers benefit the scientists who submit, part of the solution should also come from developing countries themselves. The support from funders must include provision for submission fees, so that government agencies that support research projects take responsibility for their investment. I echo the Salvador Declaration on Open Access for Developing Countries (www.icml9.org/meetings/openaccess/public/documents/declaration.htm) and I urge governments of these countries to consider the cost of publication as part of the cost of the research.

Hernán A. Burbano

Department of Biology, Universidad Nacional de Colombia, Ciudad Universitaria, Bogotá, Colombia

BOOKS & ARTS

Selling Darwin

Does it matter whether evolution has any commercial applications?

The Evolving World: Evolution in Everyday Life

by David P. Mindell

Harvard University Press: 2006. 270 pp.

£16.95, \$25

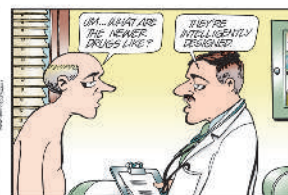
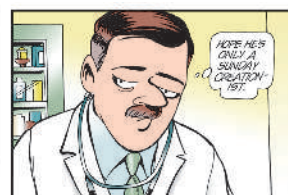
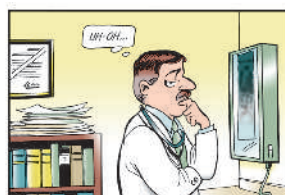
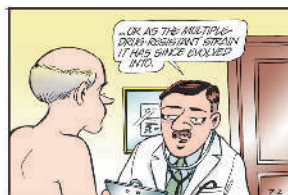
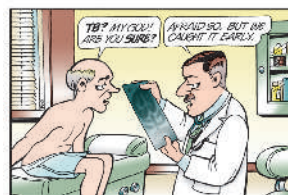
Jerry A. Coyne

After lecturing this spring to the Alaska Bar Association on the debate over intelligent design and evolution, I was approached at the podium by a young lawyer. The tight-lipped smile, close-cropped hair and maniacal gleam in his eyes told me that he was probably a creationist out for blood. I was not wrong. "Professor Coyne," he said, voice quivering with anxiety, "I don't agree with what you said about evolution, but even if it were true, how does it cash out?" "Excuse me?" I answered. "Cash out!" he said. "Does it have any practical value? What good is it?" By 'good', of course, he meant money. After a moment's thought, I muttered something about drug resistance in bacteria, adding that not all research could — or should — be about money.

The idea that the main virtue of science lies in its practical applications, especially in fighting disease or creating wealth, is a by-product of the American notion that everything comes down to the dollar. After all, in a country where Martin Luther King dreamed that people should be judged by the content of their character, they are still judged by the cost of their car. It is a peculiarly American objection to evolution that it can't cure cancer or make you rich. And some US biologists, steeped in a culture both mercenary and resistant to evolution, believe that to sell darwinism to people we must show them how darwinism helps people to sell.

This is the motivation for David Mindell's engaging book *The Evolving World*. As he notes: "When a concept and its resulting applications become useful, people tend to embrace the applications and, eventually, the underlying concepts. It is difficult to argue with success." Indeed. Other fascinating aspects of science may lack practical application (work on black holes, for instance), but these apparently don't need justification because they don't strike at the core of human values as evolution seems to do.

After reminding us that acceptance of evolution has been no slower than that of the heliocentric theory of the Solar System, or the germ



theory of disease, Mindell gets down to what he sees as the practical uses of evolutionary biology. These include plant and animal breeding, understanding the evolution of drug resistance in microbes and pesticide resistance in insects, darwinian medicine, and evolutionary conservation biology. Many of the examples of artificial breeding, although familiar, should inspire the lay reader. If we had the skeletons of only chihuahuas and St Bernards as fossils, these breeds of dog would be considered not only different species (which they may well be, given the difficulty of cross-copulation), but also members of different genera. Darwin used artificial selection in *The Origin of Species* to persuade the public of the power of natural selection, and our vastly increased knowledge of the history of human selection on animals and crops has confirmed his argument many times over.

Mindell likewise gives a readable account of evolutionary theory in medicine. Phylogenetic analysis has helped us trace the animal origins of human pathogens such as anthrax, tuberculosis, AIDS and influenza, as well as more specific routes of infection; for example, the testimony of a systematist helped convict a Louisiana doctor of injecting his mistress with HIV-infected blood. And darwinian medicine gives insight into why virulence and transmission are evolutionarily connected: malaria, reliant on the mosquito vector, leaves

its victims prostrate and susceptible to bites, whereas the common cold, spread through the air, leaves its victims free to move around. But an evolutionary viewpoint has not led to cures, so its contribution to medicine has been more heuristic than practical.

In the field of conservation, Mindell highlights the use of phylogenies to recognize and catalogue the biodiversity that can yield valuable drugs. Conservation genetics — the use of genetics to save endangered species — is given short shrift, but that seems fair given its alarmingly low rate of success so far.

Mindell's defence of evolution ends with two odd chapters: one on 'evolutionary metaphor in human culture', the other on 'the role of evolution in court and classroom'. There are broad parallels between biological evolution and the evolution of languages and religions, but little more. And his legal examples, notably forensic DNA and forensic entomology, have little to do with evolution, while speculation about the evolutionary basis of ethics is a notorious intellectual quagmire.

As a brief for the practical value of evolution, *The Evolving World* gets a mixed verdict. It is embellished with good examples, and anybody who has not been exposed to the role of evolution in human affairs will undoubtedly derive some benefit. But there are problems too. In his desire to show how useful evolution is, Mindell strives desperately to herd every stray

G. B. TRUDEAU/UNIVERSAL PRESS SYNDICATE

area of biology, even those barely related to evolution, into the darwinian fold. The “fruits of biodiversity” could yield useful compounds whether they were evolved or created. If our “evolved capacity for learning and planning” helps us solve conservation problems, it also produces art and psychotherapy. Perhaps our public-health practices “are dictated by the principles of evolutionary population genetics”, but the Romans built their aqueducts for supplying fresh water without the benefit of reading R. A. Fisher, J. B. S. Haldane and Sewall Wright.

To some extent these excesses are not Mindell's fault, for, if truth be told, evolution hasn't yielded many practical or commercial benefits. Yes, bacteria evolve drug resistance, and yes, we must take countermeasures, but beyond that there is not much to say. Evolution cannot help us predict what new vaccines to manufacture because microbes evolve unpredictably. But hasn't evolution helped guide animal and plant breeding? Not very much. Most improvement in crop plants and animals occurred long before we knew anything about evolution, and came about by people following the genetic principle of ‘like begets like’. Even now, as its practitioners admit, the field of quantitative genetics has been of little value in helping improve varieties. Future advances will almost certainly come from transgenics, which is not based on evolution at all.

As far as I know, there have been only two genuine commercial applications of evolutionary theory. One is the use of ‘directed evolution’ to produce commercial products (such as enzymes to protect crop plants from herbicides). The other is the clever use of insecticide-free ‘pest refuges’ to stop herbivorous insects evolving resistance to herbicides containing *Bacillus thuringiensis* (Bt) toxins, a strategy derived from principles of population genetics. There will certainly be more of these to come. And evolutionary algorithms are used in designing computer programs, and may have uses in engineering and economics.

One reason why Mindell might fail to sell Darwin to the critics is that his examples all involve microevolution, which most modern creationists (including advocates of intelligent design) accept. It is macroevolution — the evolutionary transitions between very different kinds of organism — that creationists claim does not occur. But in any case, few people actually oppose evolution because of its lack of practical use. As with my Alaskan interlocutor, they oppose it because they see it as undercutting moral values.

All the same, Mindell's analogy between biological evolution and the evolution of languages can be used to refute the tiresome creationist claim that we haven't seen one species change into another. We haven't seen one language change into another either, but any reasonable creationist (an oxymoron?) must accept the clear historical evidence for linguistic evolution. And we have far more

fossil species than we have fossil languages.

In the end, the true value of evolutionary biology is not practical but explanatory. It answers, in the most exquisitely simple and parsimonious way, the age-old question: “How did we get here?” It gives us our family history writ large, connecting us with every other

species, living or extinct, on Earth. It shows how everything from frogs to fleas got here via a few easily grasped biological processes. And that, after all, is quite an accomplishment. ■

Jerry A. Coyne is in the Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637, USA.

Triumph and dismal failure

Technology Matters: Questions to Live With

by David Nye

MIT Press: 2006. 282 pp. \$27.95

Don Ihde

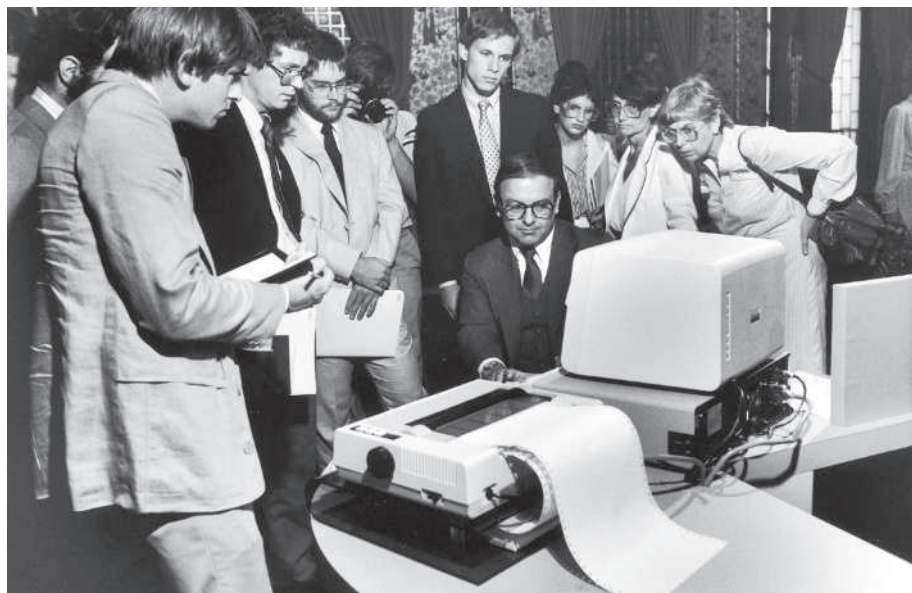
Humans and our ancestors have been using technologies since they made the first stone tools 1.3 million years ago, David Nye points out in his book *Technology Matters*. Yet the term ‘technology’ has been in widespread use for less than 100 years. A survey of prominent US periodicals published between 1860 and 1870 yields only 149 references to ‘technology’, compared with 24,957 mentions of ‘inventions’. Nye credits the Norwegian sociologist and economist Thorstein Veblen with giving ‘technology’ its more contemporary sense, and concludes that the word only gained common currency after the First World War.

Nye is a historian of technology and his book focuses on the difficult problem of showing how technologies matter. To do that requires some insight not only into the history of technologies, but into their predictability. The historian Thomas Carlyle described economics as the “dismal” science, but it would seem our history of predicting changes in technology is even worse. Nye cites the work of George Wise, a historian associated with General Electric, whose doctoral thesis revealed that of

the 1,500 published predictions from scientists, inventors and sociologists he surveyed, only a third were fulfilled. Towards the end of the book Nye claims that historians are more likely than most to get their predictions right. I found little proof of this claim in the book.

One could use this predictability problem as evidence of the indeterminacy of technologies; that is, they have multiple, but indefinite, effects. And, indirectly, Nye does this. He is clearly against the now outmoded notion of technological determination. One of the positive features of his book is the vast array of examples and mini-histories that are developed. Nye recognizes the tendency of inventors to hype each invention and make grand claims about how it will bring about a utopian future. Yet a more sober historical account shows both that the outcome may be very different to that predicted, and that much effort has to be put into getting the technology accepted. For example, it took a long time to create demand even for the technologies that shaped much of the modern world, such as the telegraph, the telephone and even the personal computer. Samuel Morse, Nye points out, spent five years “lecturing, lobbying, and negotiating” before getting the US Congress to pay for the first telegraph line.

A second theme for Nye is the claim that technologies are “socially constructed”, which



When IBM launched its early home personal computers, it had to create a demand for them.

R. MORSE/TIME & LIFE/GETTY

he means in a broad sense. He argues that any emergent technology requires many players, not just an inventor with a great idea. There is a field of players involved in negotiations and development, and only some of the resulting technologies are ultimately successful. This plays back into the problem of predictability.

What's missing from Nye's account — and this is a complaint that can be levelled at 'social constructionists' — is a sensitivity to the 'materiality' of technologies. Part of the struggle is not just that of markets, social structure and the like, but also the resistance and accommodation of the material used. For example, consider human powered flight. Most of Leonardo da Vinci's machines would have failed to fly because of their sheer weight and lack of strength. But when the Gossamer Albion was constructed in 1979 using modern

materials, kevlar and mylar, it was moderately successful. One exception for Leonardo, according to a recent account (see *Nature* 421, 792; 2003), was a glider, but this was modified by subtracting heavy control mechanisms and adding hang-glider strengtheners.

Nye's book addresses many of the issues and debates surrounding our highly textured technological society, and these are reflected in the questions he asks. Does technology control us? Does it lead to cultural uniformity or diversity? To sustainable abundance or to ecological crisis? To more security or escalating danger? The book is rich in examples, is easily readable and is short enough to be recommended for a day's read. ■

Don Ihde is in the Department of Philosophy, Stony Brook University, Stony Brook, New York 11794, USA.

about former patterns of vegetation, land use and climatic changes.

Given my own background in west Africa and work on the interaction between crops and livestock, I particularly enjoyed the chapter on African soils. It presents well the different assumptions brought to sub-Saharan Africa by colonial administrators, and their ever-ready desire to transplant foreign agricultural knowledge into local farming systems. And despite occasional massive failures with such enterprises, such as the groundnut scheme in Tanganyika in the 1940s and 50s, the power of the narrative linking African peasant farmers to poor soil management has enabled confidence among the soil experts to be sustained even today. As the book points out, "theory moulds perceptions of environmental change", made worse by the lack of detailed data on which to make firm judgements. Yet soil scientists are slowly starting to break free from conventional approaches to soil and water conservation and turn to more participatory methods that recognize local knowledge and expertise. These are now bearing a harvest of their own, even in areas with low rainfall such as the West African Sahel, with the spread of simple methods for catching rainfall and concentrating both water and organic matter along with the seed in small planting pots.

Thankfully, the past five years have seen a massive growth in demand for organic food, bringing a return to the kind of knowledge and understanding among farmers about what works best for the diverse contexts and constraints faced in different parts of the farm landscape. With luck, the rising cost of oil and energy will also help shift farm production away from high-intensity inputs and mechanization towards the greater use of agro-ecological processes. It's good to know that in this book and its detailed reference list there lies a sound body of historical material on soils that has been laid down over time. ■

Camilla Toulmin is director of the International Institute for Environment and Development, 3 Endsleigh Street, London WC1H 0DD, UK.

On fertile ground

Soils and Societies: Perspectives from Environmental History

edited by J. R. McNeill & Verena Winiwater

White Horse Press. 2006. 369 pp. £50

Camilla Toulmin

Soil is often viewed as just 'dirt', but *Soils and Societies*, edited by J. R. McNeill and Verena Winiwater, shows why soil provides the foundation on which our societies are based. Soil enters ordinary speech in many and various ways. We talk of seed falling on fertile ground, and the transformation wrought by rain falling on parched soils. Long-time urban dwellers wait for years for their name to top the list for a small allotment to grow fruit and vegetables. In Marcel Pagnol's novel *Jean de Florette*, the central characters taste the soil of a field they hope to acquire to get an indication of its quality — a practice dating back, so *Soils and Societies* tells us, to Roman times and before.

Western society today is particularly disconnected from farming life, leading to contradictory demands on what we want the food and agricultural system to do for us. So this book is a valuable reminder of the former close understanding of the soil that human societies needed to harvest higher returns and maintain soil structure and quality. A delight for the soil aficionado, this book teases out multiple threads from different societies and weaves them together to show how the fabric of daily life is closely connected to the soils from which their bread has been harvested.

Eleven chapters range from a discussion of early Indian poetry about the natural world and a history of the soils of Mesoamerica to a discussion of nutrient flows in pre-modern European agriculture. The chapter on the dynamics of soil, landscape and culture on Easter Island provides welcome clarity on how and why this

extraordinary culture, known for its massive stone sculptures of heads, came to an end.

In the introductory chapter, the editors provide a valuable framing of the main issues. Ultimately we all depend on the continued health of certain fundamental ecosystem elements and processes. As the Millennium Ecosystem Assessment reported last year, the health of our soils, in wetlands, forest areas and dry savannas, is declining. Soils have their own histories, natural and human, which intertwine, leading either to declining productivity or, if they are well managed, to sustaining the heart of the farming system.

Soils change in different ways, depending on the time scale. Gully erosion can open up a hillside in a few hours after an exceptional storm. Over decades, the barely noticeable losses each year from sheet erosion cumulatively bring about a scarred landscape, like skin drawn too tight over a bony visage. The sediment found in lake beds forms a rich archive of information



Laying the groundwork: growing crops in areas with low rainfall requires good soil management.

V. WENTZEL/HULTON ARCHIVE/GETTY

Out of the darkness

The Black Cloud

by Fred Hoyle

Heinemann/Harper and Brothers: 1957

Jay M. Pasachoff

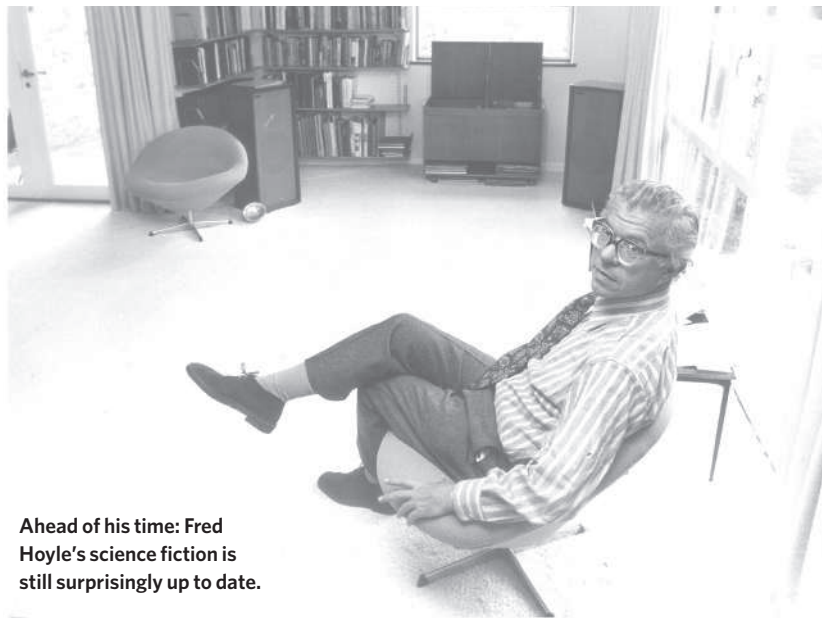
From time to time, I remember an image of a dark interstellar cloud invading the Solar System and of scientists wondering what to do. The memory is not real, of course, but conjured up from reading Fred Hoyle's novel *The Black Cloud* long ago. I was inspired to re-read it by Simon Mitton's recent excellent biography of Hoyle, *Fred Hoyle: A Life in Science* (Aurum Press, 2005), and found the story to be much more up-to-date than I had expected.

Scientists read science fiction by their peers for the scientific ideas and the characterizations of people they may know, not for the literary style. The science has held up fairly well — interstellar clouds are still important for astrophysics. What I find particularly well done in Hoyle's book is the portrayal of how scientists work, figuring out basic parameters and making first-order estimates based on inadequate data. The reasoning processes of the scientists come up in several contexts, such as how they work out when the cloud will arrive, or that the cloud is slowing down, not speeding up under gravity, as it approaches the Sun.

What I was not expecting was the applicability of political matters in the book to today's situations. In a week when a US federal judge issued an opinion forcing the authorities to release the names of prisoners at Guantanamo Bay that had long been kept secret, the book's portrayal of the US and UK governments' attempt at secrecy over the black cloud's approach seemed particularly relevant. Further, when informed that a dark cloud was approaching, the British home secretary muses: "I am quite sure ... that we can dig up some regulation that will enable us to detain the two of them, the Astronomer Royal and the man from Cambridge." The prime minister responds that "the Statue Book doesn't go back so many centuries for nothing".

A main theme of the book is the conflict between the scientists, pure in thought, and the politicians of both the United States and Britain. At one point the politicians fire nuclear rockets at the cloud, not consulting the scientists (who have contrived to sequester themselves under isolated and extremely comfortable circumstances in a country house in the Cotswolds), thereby risking the future of humanity.

Looking back in terms of today's torrent



Ahead of his time: Fred Hoyle's science fiction is still surprisingly up to date.

of digital information, it is fun to see Hoyle's speculations from the mid-1950s of information transfer rates, starting with the "human mouth transmits information at some two words per second".

One can imagine the characters as they were then — no doubt with Hoyle as the brilliant, young British scientist and Willy Fowler of Caltech as the Pasadena genius. The astronomer royal of the time has been superseded in my mind by today's Martin Rees, just the brilliant theoretician needed to save the world. The original need for secrecy in the 1957 novel proved prescient for the withholding of information about the discovery of pulsars in 1968 for some weeks, when it was thought that Jocelyn Bell's bits of "scruff" might be signals from little green men. Since then, a protocol has been set up for how scientists should act and release information in case a believable signal is received from one of the projects in the search for extraterrestrial intelligence.

The rivalry among Palomar Observatory scientists, British theoreticians, and British and Australian radioastronomers has certainly evolved, but the basic pros and cons of their cooperation rings true.

Hoyle dwells on the relations between scientists and those in other fields. He was writing at the same time that C. P. Snow was formulating his lectures on 'The Two Cultures and the Scientific Revolution'. Hoyle's scientist may have quoted (actually, slightly misquoted, perhaps on purpose) Shakespeare's "Come kiss me, sweet and twenty/Life's a stuff will not endure"

from *Twelfth Night* but we don't find non-scientists quoting, as Snow demanded, the second law of thermodynamics or its equivalent.

Iconoclastic Hoyle, of course, interspersed a variety of ideas through the book. For example, he pointed out the need for dense urban planning to survive the cold that would be caused when the cloud masked the sun. And I am sure he couldn't resist sticking in some backing for his steady-state theory: "'I would not agree that there ever was a first member', said the Cloud," speaking of its ancestry. "'Oh-ho, there we go. That's one in the eye for the exploding-universe boys'."

Students and the general public could learn some valid statistical ideas from the way the book's scientists analysed the incoming data: "It's no good doing a lot of experiments first and then discovering a lot of correlations afterwards ... Otherwise it's like betting on a race after it's been done." Would that so much pseudoscience and alternative medicine were analysed in this way.

Hoyle may have pushed his speculations — including those about interstellar matter — too far in his later years, but his fertile mind came up with some lasting ideas as well, such as the one about energy levels in the carbon atom that led to our understanding of how most of the chemical elements form. *The Black Cloud*, on a pleasant but less significant level, remains another testimony to his skills.

Jay M. Pasachoff is Field memorial professor of astronomy, Williams College, Williamstown, Massachusetts 01267, USA.

IMMUNOLOGY

Protection and privilege

Herman Waldmann

The immune system not only attacks microbes, but also regulates itself to avoid harming vital organs. Cells notorious for their involvement in allergy turn out to be vital to this protective function.

The immune system's capacity to regulate its own activities can be exploited to prevent rejection of transplanted organs and to reverse autoimmune diseases such as diabetes. The chief agents of this immune self-regulation system are specialized cells — T regulatory cells (T_{reg}) — that somehow prevent other immune cells from attacking healthy tissues. In this issue (page 997)¹, Lu *et al.* demonstrate that the protective actions of T_{reg} cells depend on so-called mast cells, which were hitherto bad-mouthed for their undesirable role in allergic responses*. Perhaps, at last, a virtuous physiological role for mast cells has been uncovered.

Long-term tolerance of transplanted tissues by the body can be achieved by using a short treatment of antibodies that block vital co-receptors (CD4 and CD8) or co-stimulatory molecules (CD154) on a subset of immune cells (T cells). This reprogramming of the immune system depends on the activity of specialized regulatory $CD4^+$ T cells (T_{reg}) that are drawn to the transplanted tissue and its surrounding milieu. The T_{reg} cells confer on the tissue some special exemption from attack, just as the fetus in the womb is protected from the mother's immune system. A major goal of current immunological research is to establish just how T_{reg} cells ensure this ceasefire.

A clue to what that mechanism is came from the observation that cultures of T_{reg} cells tend to be contaminated with 'unwanted' mast cells². This was attributed to a secreted molecule called IL-9 that is made by the T_{reg} cells and that enhances mast cell growth and functionality. More surprisingly, when recipient mice that were tolerating a skin graft were given a second graft from an identical donor, the second graft became infiltrated not only with T_{reg} cells, but also with mast cells³. This led to the proposal that T_{reg} and mast cells might form a functional unit that mediates graft tolerance^{2,3}.

Lu *et al.*¹ confirm these findings and extend them by performing crucial work using an experimental mouse strain (C57BL/6-Kit^{W-sh}/Kit^{W-sh}) that is genetically engineered to

*This article and the paper concerned¹ were published online on 20 August 2006.

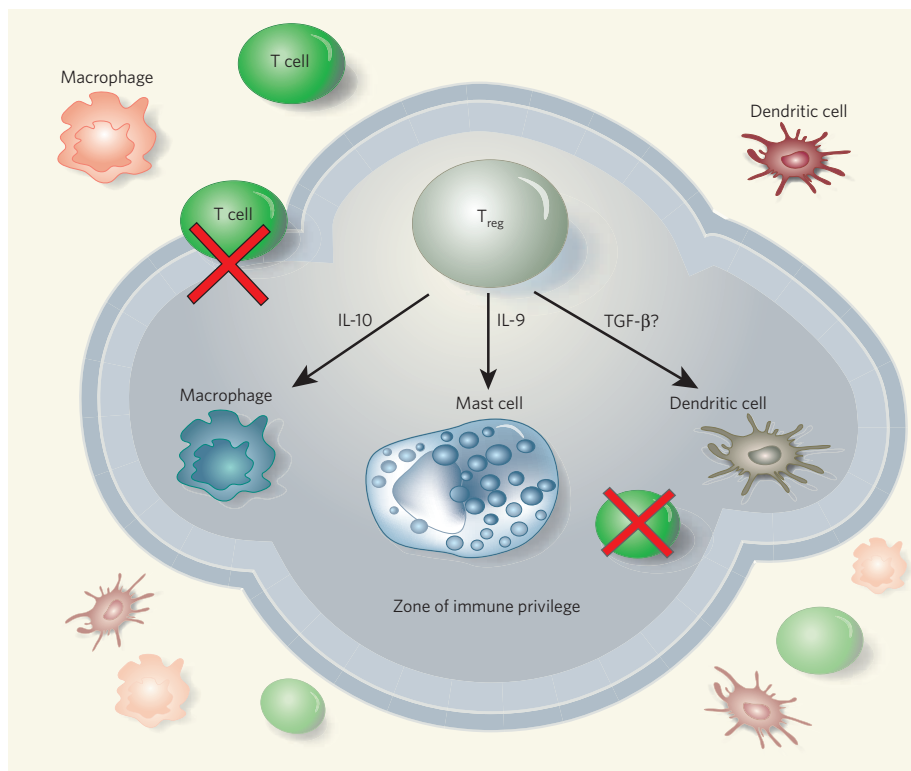


Figure 1 | Immune self-regulation: a possible mechanism. The work of Lu *et al.*¹ suggests that mast cells interact with regulatory T cells (T_{reg}) to exempt transplanted tissues from an immune response. If sufficient T_{reg} cells can enter a transplanted tissue, they could then mobilize mast cells to enhance the capacity of the mast cells to decommission immune responses within a limited 'privileged' area. By interacting with antigen in the tissue, the T_{reg} cells are activated to produce the IL-9 polypeptide and perhaps other molecules (maybe IL-10 or TGF- β). These molecules enable T_{reg} cells to interact with cells of the innate immune system (such as dendritic cells or macrophages), leading to further inhibition of immune responses. The end result is that T cells, which might otherwise kill other cells and damage tissues, are prevented from doing so and are then inactivated.

be deficient in mast cells⁴. The authors used these mice as recipients for transplants, but their attempts to induce therapeutic tolerance in the mice were unsuccessful, as all the test grafts were promptly rejected. Spectacularly, however, injecting mast-cell-enriched cell cultures into the recipient mice restored the host's capacity to accept grafts for prolonged periods.

The second major breakthrough by Lu *et al.*¹ comes from the discoveries that all categories of T_{reg} cell make IL-9, that IL-9 can be found in the

tolerated graft, and that neutralization of IL-9 *in vivo* prevents regulation of graft rejection by T_{reg} cells. The authors interpret the findings as indicating that IL-9 from T_{reg} cells is involved in the recruitment and activation of mast cells at the graft site.

Mast cells belong to the 'innate' immune system — the system that provides immediate defence against microbes that the body hasn't encountered before. Although they participate in immune responses, they do not intrinsically recognize unique antigens. The recognition of

specific antigens is mediated instead by the 'adaptive' immune system, where cells such as T cells remember previous microbe encounters and react strongly against the aliens if the infection is repeated.

When it comes to immunity towards microbes, cells of the innate immune system, including mast cells, macrophages and dendritic cells, interact with the adaptive system in two ways. First, they sense dangerous microbes through generic receptors for potential pathogens, and, once activated, they alert the adaptive system to the danger. Mast cells, for instance, expel packets of mediator molecules (including histamine) that cause an immediate inflammatory reaction and attract adaptive cells. Second, innate immune cells can act as the executive arm of the adaptive system, once armed with antibodies or mobilized to killer-mode by T-cell-derived mediators.

Recent research on macrophages and dendritic cells has taught us that the innate immune system may also mediate the opposing process of permanently decommissioning adaptive cells so that they cannot mount an immune response. For example, immature dendritic cells display antigens in a way that makes T cells tolerate that antigen so they no longer respond to it. Moreover, T_{reg} cells can manifest their suppressive activity by modulating dendritic cells, so that the dendritic cells are no longer able to alert the adaptive cells to danger.

Perhaps this symmetry — in both activating and inhibiting adaptive reactions — is a general feature of innate cells. We have tended to think of these cells as agents of destruction, waiting around until the microbe appears. Now, there are reasonable grounds for elevating their status to cells that have a positive physiological role — ensuring self-tolerance and/or tissue integrity through the maintenance of 'privileged' microenvironments where adaptive immune responses are damped down (Fig. 1).

The implications of the Lu *et al.*¹ paper go beyond transplantation. This work could form the basis for understanding why mast cells are located in very specific sites within tissues (for example, nerves, vessels, hair follicles or epithelia). It was always hard to see how this positioning was related to their potential for immune function. Perhaps these sites require some low-level immune privilege. Also, is it possible that the mast cells found within tumours contribute some immune privilege? In support of this idea, mast-cell-deficient mice seem to have some natural resistance to the induction of tumours⁵.

Which checkpoint of the immune self-regulatory response needs the mast cells is not yet established. Are they required to induce T cells to a regulatory role or are they required at some later phase in the execution of that function? Either way, the current findings will galvanize research into the molecular basis by which mast cells decommission immune

function, and provide a possible new direction for acquiring drug targets.

Herman Waldmann is in the Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK.
e-mail: herman.waldmann@path.ox.ac.uk

1. Lu, L.-F. *et al.* *Nature* **442**, 997–1002 (2006).
2. Cobbold, S. P. *et al.* *Immunol. Rev.* **196**, 109–124 (2003).
3. Zelenika, D. *et al.* *Immunol. Rev.* **182**, 164–179 (2001).
4. Galli, S. J., Nakae, S. & Tsai, M. *Nature Immunol.* **6**, 135–142 (2005).
5. Wedemeyer, J. & Galli, S. J. *Lab. Invest.* **85**, 388–396 (2005).

COSMOLOGY

Unique, or not unique?

Martin Bojowald

That is the question. The search for a single theory of everything is as old as science itself, and is now the beat of quantum cosmologists. But some basic tenets that inform the quest are being challenged.

*Whether 'tis nobler in the mind to suffer
The slings and arrows of outrageous fortune,
Or to take arms against a sea of troubles,
And, by opposing, end them?*

Hamlet Act III, Scene i

Hamlet's existential agony has always been, in a variant form, part of the mindset of those researching quantum gravity. In this field, a large faction has deemed itself to be in the outrageously fortunate position of being close to a unique theory of how the Universe is as it is. This theory would not only put descriptions of all forms of matter and their interactions on the same footing, but also reconcile the two seemingly inharmonious pillars of modern physics: general relativity, which describes gravity, and quantum theory. The belief in such an all-encompassing theory has been the driving force for the various models known as string theories, grouped under the umbrella

of 'M-theory', that have been developed in the past decades.

But there is an alternative concept. This holds that solutions, rather than models, are unique. Solutions are generally more important for physics than theories: observations are compared with properties that emerge from a theory, not with the construct itself. But finding realistic solutions to theories — solutions that reproduce the features of the Universe that we observe now — has proved much harder and messier than constructing the theories themselves.

That applies even with a modest interpretation of 'realistic', such as merely requiring the solution to undergo the kind of accelerated late expansion that our Universe appears to be going through now. Once such solutions were finally uncovered¹, myriads of them turned up in various corners of the string landscape². Thus, as there is currently no selection criterion

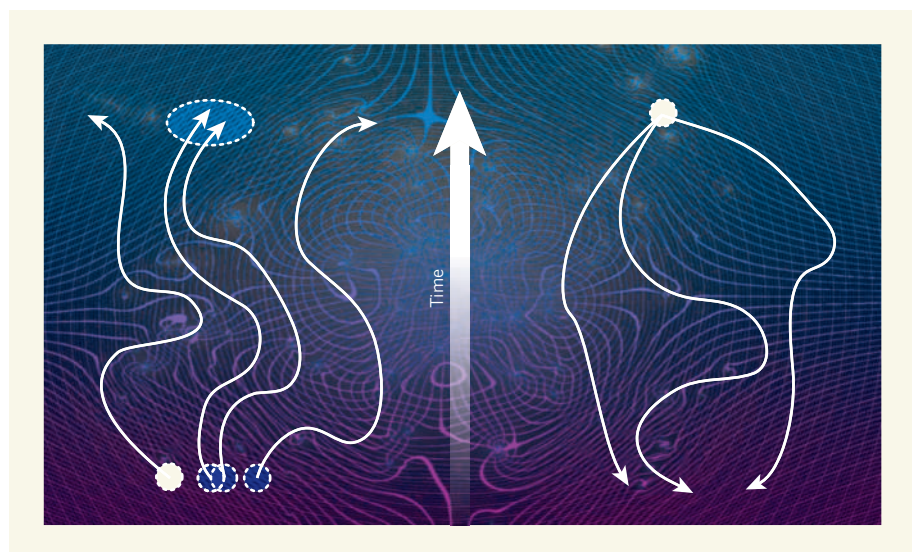


Figure 1 | Take it from the top. **a**, A bottom-up description requires fine-tuned initial values for individual histories to arrive close to a universe as we see it now. **b**, A top-down formulation avoids that problem by starting with present properties and working backwards. Both viewpoints can be used in classical and quantum physical frameworks, but the top-down interpretation is most strongly motivated by quantum cosmology. The quantum top-down approach picks suitable histories from a quantum superposition of all possible histories that lead to the current Universe. Thus here, the solution, rather than the theory that leads to it, is unique.

by which to choose among this vast range of solutions, it does not seem particularly useful to claim that any one string theory is unique.

Stephen Hawking and Thomas Hertog³, writing in *Physical Review D*, now propose arms to be taken against this sea of troubles. The arms they propose are admittedly not new, having been developed⁴ in the 1980s by Hawking, James Hartle and others. Then, too, the motivation was uniqueness of solutions: specifically, if a quantum-cosmological model explains properties of our Universe (of which, by definition, we see only one), then it should also explain why this, and only this, solution emerges. Such a model can be compared with observations, making the whole framework testable and thus predictive.

Hartle and Hawking called their condition for establishing uniqueness⁴ the no-boundary proposal, because it removed the boundaries of space-time. According to this view, the Universe is a closed surface — rather like a surface of an inflating balloon — and has no beginning in time. Such a closure of space-time is not meaningful in classical general relativity, and thus requires the introduction of aspects of quantum theory. This in turn serves to establish uniqueness: although there are several possible closures, quantum theory can, unlike classical theory, deal with all of them at once as a probabilistic superposition. Alternatives to the Hartle–Hawking proposals include Alexander Vilenkin's proposal that the Universe initially tunnels out of a quantum state where space and time are not defined⁵, and more recent models based on new formulations of quantum gravity.

In their recent paper³, Hawking and Hertog refresh the no-boundary proposal, adding new insight and giving it a new name: top-down cosmology. Looking at a space-time diagram where time runs in the upward direction, the conventional approach to cosmology is 'bottom-up' (Fig. 1a): one starts with initial conditions in the past and calculates forward to aim at properties seen now. This process usually requires very specific, fine-tuned initial values. The top-down approach (Fig. 1b) avoids this problem by taking the properties of the Universe as it appears now and calculating its history backwards. This process is applied to a quantum superposition of different Universe states, with 'final', rather than initial, conditions being set to select one history in the superposition relevant for our observations. In this way, the non-intuitive quantum superposition is reduced to a classical Universe as we observe it.

Traditional bottom-up cosmologies also suffer from the problem that they break down at points where infinite energies arise in solutions of the equations of general relativity. These points are known as singularities, and our Universe may have experienced one at the Big Bang. Starting from a simple initial state that explains the emergence of the Universe, the chances are that one will run into a

singularity before even getting close to the present. On its own, the top-down approach is not free from this problem, as the probability of hitting a singularity when calculating backwards is just as great as when calculating forwards. But when combined with the no-boundary proposal, top-down is safer: this combination has the effect of closing off singularities from classical space-time before the history being traced can approach them. Again, this act of closing off introduces aspects of quantum theory, and leads directly to a quantum description of gravity.

The arguments that Hawking and Hertog present are not complete, as they distinguish only between 'classical bottom-up' and 'quantum top-down'. That mixes up the singularity problem, which is a matter of classical against quantum theory, with the issue of predictivity. This second point amounts to what preconditions, whether initial or final, we may set when evaluating a theory and is the dividing line between bottom-up and top-down theories. Elsewhere in physics, it is clear which approach is better: one predicts final observations from the initial set-up of an experiment. But this option is clearly not available in cosmology, as we have no influence over the initial conditions of the Universe. Hawking and Hertog's suggestion is indeed a radical shift of approach: it is, as befits the description of the evolution of the Universe, much more akin to the holistic methods of 'universal history' (advocated in an early form by Friedrich Schiller in his inaugural lecture at Jena in Germany in 1789)⁶ than anything familiar from the physical sciences.

A further imprecision is that the authors sometimes use 'top-down' and 'no-boundary' interchangeably, although these are different concepts. That undermines some claims and disregards alternatives: there are, for instance, more general solutions of the singularity problem that do not require a top-down approach; and theories of decoherence^{7–10} provide detailed descriptions of the quantum–classical transition as a physical process in which a superposition evolves into a semi-classical history.

Hawking and Hertog present examples³ for final conditions that can be chosen as the starting point for the backwards computation of the Universe's history. Currently, that choice is wide open, and no clear line is drawn between anthropic conditions — conditions that must be so, because otherwise we humans could not be there to observe them — and conditions that arose accidentally during the development of the Universe, but are nonetheless regarded as important for the purposes of the computation. The fewer final conditions there are, the more predictive a theory will be. What is considered an accident or not is often just theoretical prejudice. Not distinguishing between 'accidental' and other conditions in the determining final set allows one to escape a firm decision, but could undermine the top-down approach by relinquishing deeper explanations. Indeed,



50 YEARS AGO

'The neutrino' — While careful reasoning from experimental evidence gathered about all terms in the beta-decay process... may support the inference that a neutrino exists, its reality can only be demonstrated conclusively by a direct observation of the neutrino itself... Such an experiment is made possible by the availability of high beta-decay rates of fission fragments in multi-megawatt reactors and advances in detection techniques. An estimate of the neutrino flux available from large reactors shows that a few protons should undergo reaction in 50 litres of water placed near the reactor... The complete detector consisted of a 'club sandwich' arrangement employing two target tanks between three detector tanks... located deep underground near one of the Savannah River Plant production reactors of the United States Energy Commission... After running for 1,371 hr., including both reactor-up and reactor-down time... a signal dependent upon reactor-power, 2.88 ± 0.22 counts/hr. in agreement with the predicted cross-section (6×10^{-44} cm.²) was measured... Frederick Reines and Clyde L. Cowan, jun.

From *Nature* 1 September 1956

100 YEARS AGO

'Thermodynamic reasoning' — In the address delivered by Principal Griffiths at York... I read: "Prof. Armstrong remarks that it is unfair to 'cloak the inquiry by restricting it to thermodynamic reasoning'... He adds that such a course may satisfy the physicist but 'is repulsive to the chemist'."

This statement shows a strange misapprehension of my position... At present, progress is not a little hampered by the fact that chemists and physicists cannot wander through the museums of nature in complete sympathy with one another... a confusion of language has arisen which keeps us apart: we must both strive to speak a simpler language.

From *Nature* 30 August 1906

50 & 100 YEARS AGO

without any final conditions whatsoever, we are back to a decoherence approach as most predictive: this approach attempts to describe the emergence of the current state of the Universe through a physical process in which accidents may arise, but do not affect the overall theory.

So will the no-boundary, top-down cosmology really turn the string landscape into a goldmine of physical predictions? Hawking and Hertog's paper is mainly about how to interpret physics from the top-down perspective, with few supporting calculations, so their answer remains uncertain. Clearly, with too much leeway in choosing final conditions (these might include, the authors propose, the number of dimensions in space-time, or the observed features of the standard model of particle physics), physics is in danger of becoming a tautology — a proposition already true by definition. But approached carefully, a top-down viewpoint on cosmology can, at some expense of losing explanatory power, serve well as an interpretational framework to test theories in the string landscape.

Hawking and Hertog's work also represents a

welcome attempt to combine pivotal ideas from different approaches to quantum gravity. Such cross-fertilization has rarely happened, but can only improve our overall understanding. That such efforts should continue is indeed, to return to Hamlet, "a consummation devoutly to be wished".

Martin Bojowald is in the Department of Physics, Penn State University, 104 Davey Lab, University Park, Pennsylvania 16802-6300, USA.
e-mail: bojowald@gravity.psu.edu

1. Kachru, S., Kallosh, R., Linde, A. & Trivedi, S. P. *Phys. Rev. D* **68**, 046005 (2003).
2. Susskind, L. preprint available at www.arxiv.org/hep-th/0302219 (2003).
3. Hawking, S. W. & Hertog, T. *Phys. Rev. D* **73**, 123527 (2006).
4. Hartle, J. B. & Hawking, S. W. *Phys. Rev. D* **28**, 2960-2975 (1983).
5. Vilenkin, A. *Phys. Rev. D* **30**, 509-511 (1984).
6. Schiller, F. reprinted in *Hist. Theor.* **11**, 321-334 (1972).
7. Griffiths, R. J. *Stat. Phys.* **36**, 219-272 (1984).
8. Omnès, R. J. *Stat. Phys.* **53**, 893-932 (1988).
9. Gell-Mann, M. & Hartle, J. B. *Phys. Rev. D* **47**, 3345-3382 (1993).
10. Joos, E. et al. *Decoherence and the Appearance of a Classical World in Quantum Theory* (Springer, Berlin, 2003).

SPECTROSCOPY

Shifting light with spin

Warren S. Warren

NMR spectroscopy has changed enormously over the years, but signal detection has stayed the same since the technique was invented. The latest thinking literally shines a new light on things.

Sixty years ago, the forefront of speculative physics research included the nascent field of nuclear magnetic resonance (NMR). Physicists placed atomic nuclei in a strong, constant magnetic field and then watched the voltages induced in a coil when the nuclei were perturbed with weak, finely tuned radio waves. At that time, NMR was a prime example of blue-skies research with no conceivable real-world applications. Since then, spectacular developments have established NMR as the foremost spectroscopic method for chemists, and Nobel prizes in physics, chemistry and medicine have been awarded for advances in NMR. Its descendant, magnetic resonance imaging (MRI) has become a mainstream diagnostic tool in medicine, and functional MRI — essentially, watching people think — may revolutionize neuroscience.

As part of this evolution, almost everything has changed in the basic NMR set-up, except for the method of detection; nearly all NMR and MRI experiments still use the same 'nuclear induction' concept that was used 60 years ago to detect a signal. On page 1021 of this issue¹, Savukov *et al.* report a radically different method to detect NMR signals in liquids: watching the small phase-shifts induced in a laser beam by nuclear spins.

Bringing the power of modern optics to NMR detection could greatly improve image resolution, and perhaps even sensitivity. This would be a huge step forward, as many applications, for example in the emerging field of molecular imaging, are limited by these issues.

This is not the first attempt to find a different detection method for magnetic resonance. For example, a tiny magnet placed close to a particle experiences a force exerted by the spin of that particle. Several groups have tried to detect this force directly. Signals from single electron spins² or from a few thousand nuclear spins³ have been observed in this way, but the method is not suitable for macroscopic samples.

There have even been other attempts to use optical detection in magnetic resonance. Under certain circumstances, electron spin-flips are coupled to changes in electronic energy levels, so that electronic spins can simply be measured by observing the light absorbed or emitted as the electrons jump between these levels^{4,5}. More generally, there have been many attempts to look at variations in the NMR signals of liquids caused by the optical irradiation of those liquids. The largest changes are expected in the frequencies of signals from electronically excited molecules. Such changes are predicted to be sizeable, but the process required to create

an excited population of molecules, known as near-resonant excitation, also heats the sample. This heating induces frequency shifts of its own⁶ and severely limits the utility of the method.

The technique now reported by Savukov *et al.*¹ is known as nuclear-spin optical rotation (NSOR). This method looks for phase-shifts induced in a laser beam as it passes through a liquid, rather than for frequency shifts of signals in an NMR spectrum. Detecting an optical effect instead of nuclear spins has several advantages. Measurements of NMR frequency shifts require uniform laser irradiation of the sample, but NSOR can work with much smaller, tightly focused laser beams, which in principle permits micrometre-resolution measurements. This would be a vast improvement over existing techniques; obtaining even 100-micrometre resolution in MRI is challenging, for example.

The NSOR method of detecting phase-shifts in a laser beam allows experimental designs that are less sensitive to sample heating compared with previous methods. Even three-dimensional tissue imaging is not out of the question with NSOR, as near-infrared light can penetrate many centimetres into tissue. NSOR is also enhanced by so-called hyperfine effects, which increase with the mass of the nucleus under investigation. This makes the technique particularly suitable for heavy nuclei, which generally give poor spectra in traditional NMR experiments.

The greatest problem for NSOR at the moment is its sensitivity, which is not yet as good as conventional NMR. But optical detectors can be highly efficient, as single photons are much easier to detect than single spins, and Savukov and colleagues¹ suggest several potential improvements to their technique — for example, bouncing light off mirrors to pass it through the sample many times. In addition, modern technology gives us incredible control of the timing, shape and polarization of ultrafast laser pulses. It is easy to see how NSOR, with a combination of tailored light and modulated magnetization, could couple modern optical imaging methods to MRI. If this happens, it will be yet another reminder that speculative physics research can yield great practical dividends.

Warren S. Warren is at the Duke Center for Molecular and Biomolecular Imaging, Box 90346, Durham, North Carolina 27708, USA and at Duke University, Durham, North Carolina 27708, USA.
e-mail: warren.warren@duke.edu

1. Savukov, I. M., Lee, S.-K. & Romalis, M. V. *Nature* **442**, 1021-1024 (2006).
2. Rugar, D., Budakian, R., Mamin, H. J. & Chui, B. W. *Nature* **430**, 329-332 (2004).
3. Mamin, H. J., Budakian, R., Chui, B. W. & Rugar, D. *Phys. Rev. B* **72**, 024413 (2005).
4. Jelezko, F., Gaebel, T., Popa, I., Gruber, A. & Wrachtrup, J. *Phys. Rev. Lett.* **92**, 076401 (2004).
5. Kohler, J. et al. *Nature* **363**, 242-244 (1993).
6. Warren, W. S., Goswami, D. & Mayr, S. *Mol. Phys.* **93**, 371-375 (1998).

CELL BIOLOGY

Taking a turn into the nucleus

Ulrike Kutay and Petra Mühlhäusser

How soluble proteins get into the cell nucleus is known in great detail, but how membrane proteins make it into the inner nuclear membrane has long been an enigma. The two processes in fact turn out to be related.

Eukaryotic cells house their genome in a dedicated intracellular compartment — the cell nucleus. The boundary of this organelle is a double lipid bilayer termed the nuclear envelope, which consists of an outer nuclear membrane (ONM) that is continuous with the membrane system of the endoplasmic reticulum (ER), and an inner nuclear membrane (INM) of a different protein composition. Proteins of the INM play a pivotal role in genome organization. They are initially inserted into the ER, but how do they get to their final destination in the INM? In this issue, Günter Blobel and colleagues (King *et al.*, page 1003)¹ address this long-standing question*. They report a receptor-mediated transport mechanism by which transmembrane proteins are targeted to the INM in yeast.

Exchange of material between the nucleus and the cytoplasm occurs through pores where the

*This article and the paper concerned¹ were published online on 23 August 2006.

ONM and INM are interconnected (Fig. 1). These pores are occupied by nuclear pore complexes (NPCs) that act as molecular sieves and serve as gateways for the active, bidirectional transport of macromolecules. Molecules of relative molecular mass (M_r) less than 40,000 can passively diffuse through the NPC. Transport of larger macromolecules depends on nuclear targeting signals that are recognized by shuttling transport receptors that facilitate passage through the central NPC channel. Although a variety of signals and receptors for soluble nuclear proteins have been identified, mechanisms governing the passage of INM proteins through the NPC have remained largely elusive.

After insertion into the membrane, INM proteins move by lateral diffusion through the lipid bilayer of the ER and ONM². To reach the nuclear interior, they take a hairpin turn to the INM along the bend made by the pore membrane. The simplest explanation of their route to the INM is the diffusion–retention model.

This posits that, upon arrival, association with nuclear components tethers the proteins, retaining them in the INM^{3,4}. But free diffusion along the pore membrane might well be hindered by NPC subcomplexes that surround and inhabit the pore membrane. Indeed, transport to the INM requires energy, so the diffusion–retention model might be too simplistic⁵. It may be that active remodelling of the NPC or a receptor-mediated mechanism is required for passage along the pore membrane.

Blobel and colleagues¹ now describe how transmembrane proteins are targeted to the INM in budding yeast. The authors studied the transport of two novel INM proteins, Heh1p and Heh2p (relatives of known mammalian INM proteins), which they identified from database searches. Detailed sequence analysis of yeast and mammalian INM proteins, including Heh1/2p, revealed elements resembling the classical nuclear localization signal (NLS). These amino-acid sequences occur in a subset of soluble nuclear proteins that are imported by a transport receptor called kap- α / β 1 (made of karyopherin- α and karyopherin- β 1).

The identification of NLS-like sequences in INM proteins raised the intriguing idea that transport receptors for soluble proteins might also facilitate the passage of membrane proteins through the NPC. Indeed, this is exactly what Blobel and colleagues found. First, mutations in components of the RanGTPase system, which determines the direction in which proteins are transported, impaired Heh protein import. Second, proper INM targeting of Heh proteins specifically required kap- α / β 1.

Additional experiments confirmed that the NLS-like sequence in Heh2p acts in INM targeting: it interacted directly with kap- α ; it caused nuclear accumulation when attached to a protein normally found in the cytoplasm; and its mutation resulted in Heh2p mislocalization. Remarkably, Blobel and colleagues found that the targeting of Heh2p to the INM is uniquely supported by kap- α / β 1, as the replacement of the Heh2p NLS with nuclear targeting signals specific for other import receptors failed to sustain correct localization. A receptor-specific path might therefore exist for funnelling INM proteins along the pore membrane, through the outer regions of the NPC.

Might such karyopherin-mediated protein sorting to the INM be conserved between yeast and higher eukaryotes? NLSs are predicted in a variety of mammalian INM proteins¹, but their functional relevance has yet to be explored. Interestingly, a shortened kap- α variant, importin- α -16, interacts with amino acids close to the transmembrane domain of a baculovirus-derived INM protein⁶. Importin- α -16 cannot associate with kap- β 1, however, and the involvement of importin- α -16 in the passage of INM proteins across the NPC needs to be proved.

It remains to be seen whether all natural INM proteins rely on receptor-mediated translocation. Analysis of an artificial protein chimera bearing a small nucleoplasmic domain

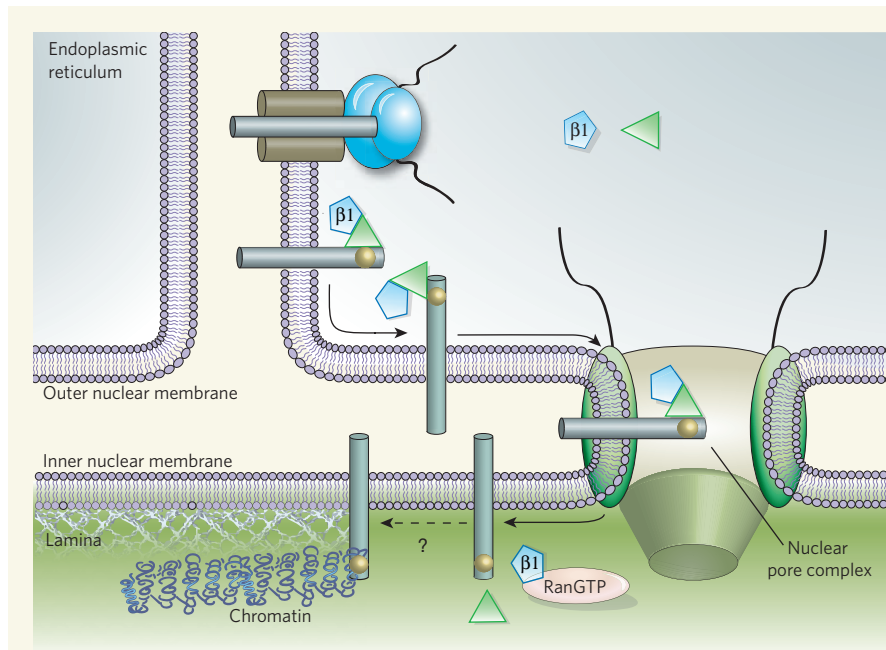


Figure 1 | Receptor-mediated transport of proteins to the inner nuclear membrane (INM). INM proteins are initially inserted into the endoplasmic reticulum membrane. Blobel and colleagues¹ show that INM proteins bearing a nuclear localization signal (gold circle) in their cytoplasmic domain are then bound by a complex of two karyopherin proteins. Karyopherin- α (triangle) recognizes the nuclear localization signal, and karyopherin- β 1 (pentagon) probably mediates the interactions with the nuclear pore complex to drive translocation of the INM protein to the nuclear side. In the nucleus, binding of RanGTP to β 1 leads to the dissociation of the import complex. The INM protein may then be retained in the nucleus by interactions with chromatin in yeast (and/or the nuclear lamina in higher eukaryotes).

of M_r 12,000, but lacking an NLS, showed that NLSs are not absolutely required for INM targeting⁵. Surprisingly, transport of this chimera is energy dependent, but which step requires energy is unclear.

It may be that receptor-mediated transport is essential for two types of INM protein: those with domains of size, folding or charge that prevents passive diffusion; and those that are not retained stably in the nucleus. In higher eukaryotes, many INM proteins are retained by interaction with chromatin (DNA and associated proteins) and/or with the nuclear lamina, a filamentous network that underlies the INM. As yeast cells have no nuclear lamina, it will be interesting to see whether nuclear retention of

yeast INM proteins is as important for INM targeting as it is in higher eukaryotes. ■
Ulrike Kutay and Petra Mühhlhäusser are at the Institute of Biochemistry, ETH Zurich, Schafmattstrasse 18, 8093 Zurich, Switzerland. e-mails: ulrike.kutay@bc.biol.ethz.ch; petra.muehlhaeuser@bc.biol.ethz.ch

1. King, M. C., Lusk, C. P. & Blobel, G. *Nature* **442**, 1003–1007 (2006).
2. Ellenberg, J. *et al.* *J. Cell Biol.* **138**, 1193–1206 (1997).
3. Smith, S. & Blobel, G. *J. Cell Biol.* **120**, 631–637 (1993).
4. Soullam, B. & Worman, H. J. *J. Cell Biol.* **120**, 1093–1100 (1993).
5. Ohba, T., Schirmer, E. C., Nishimoto, T. & Gerace, L. *J. Cell Biol.* **167**, 1051–1062 (2004).
6. Saksena, S. *et al.* *Nature Struct. Mol. Biol.* **13**, 500–508 (2006).

ASTROPHYSICS

Shock breakout caught on camera

Timothy R. Young

What exactly is the relationship between bursts of cosmic γ -rays and the stellar explosions known as supernovae? Intimate, it seems: highly magnetic neutron stars might even have spawned both.

Four papers in this issue^{1–4} present the first sightings of a remarkable cosmic event: the evolution of a hugely energetic γ -ray burst into a fully fledged stellar explosion — a supernova. It's the first time these two phenomena have been observed with the same telescope, NASA's satellite-based Swift telescope, and the implication of a common origin for both is intriguing.

Supernovae most commonly occur when a mature star ceases to generate enough energy from thermonuclear fusion to support its own gravity. A catastrophic, explosive collapse ensues, during which material from overlying stellar layers falls inwards. This creates a shockwave that rebounds outwards, fuelled by energy gained probably from internal magnetic fields and rotation. At 'shock breakout', when the shockwave emerges from the surface of the collapsing star, its energy is unleashed. It is sent out into space as radiation of all frequencies over a period of days to months — the classic signal of a supernova.

The idea that the comparatively short, sharp blast of a γ -ray burst (GRB) is perhaps some form of supernova early-warning signal had been around for some time, and over the past seven years, three candidate GRB-supernova pairings have been identified^{5–8}. But no one had ever obtained the clinching evidence of a connection: witnessing the evolution of a γ -ray burst to the all-frequency display of a supernova. In fact, no supernova had ever been observed at the moment of shock breakout.

That changes now, with observations at several wavelengths^{1–4} of an object — variously known as supernova SN 2006aj and γ -ray burst GRB060218 — that flared up on

18 February 2006. All four papers make clear that the exploding object sent out both a slightly aspherical shockwave, typical of a supernova, and a jet-like stream of material characteristic of a GRB.

Using, respectively, X-ray data and optical light curves, Campana *et al.* (page 1008)¹ and Pian *et al.* (page 1011)² show that the star concerned is a 'Wolf-Rayet' star that exploded while in a compact state, in which it contained no hydrogen or helium. That identification is supported by computer modelling presented by

Mazzali *et al.* (page 1018)³. Estimates, based on the X-ray data, of how a shell of gas at 2 million degrees expands constrained the radius of the progenitor to 1.2×10^7 km, much smaller than typical exploding stars¹. The optical light curve and spectra are characteristic of the explosion of a bare carbon-oxygen stellar core².

According to Campana *et al.*¹, the X-ray spectra of the event exhibit two distinct components: a slightly off-spherical thermal component that is most probably explained by the heating effect of the supernova shockwave as it exited the star; and a highly directed non-thermal X-ray jet. The authors think that this jet can be ascribed to the classic GRB mechanism — radiation given off by highly relativistic gas, which travels at speeds only a thousandth of a percent lower than that of light. That emission occurs most probably along the axis of rotation of an exploding star.

Pian *et al.*² and Soderberg *et al.* (page 1014)⁴, on the other hand, observe that the peak flux of GRB060218 occurred at an energy of 5 kiloelectronvolts (keV), at the lower-energy end of the X-ray spectrum. Classic energetic GRBs have a peak in their emissions at an energy of around 250 keV. Thus, the observed event would be classified as a lower-energy 'X-ray flash', and might be incompatible with the classic GRB model.

Interestingly, Soderberg *et al.*⁴ also find in their radio-frequency data the signature not of a highly relativistic jet, but of mildly relativistic expanding debris, travelling at just 90% the speed of light. This, again, is presumably a signature of the supernova shockwave. Although the jet energy from the X-ray data¹ indicated that GRB060218 was fainter overall than normal, the energy in the shockwave⁴ seems to imply that the supernova was brighter than usual. If the event had been further away from us — it is in fact the

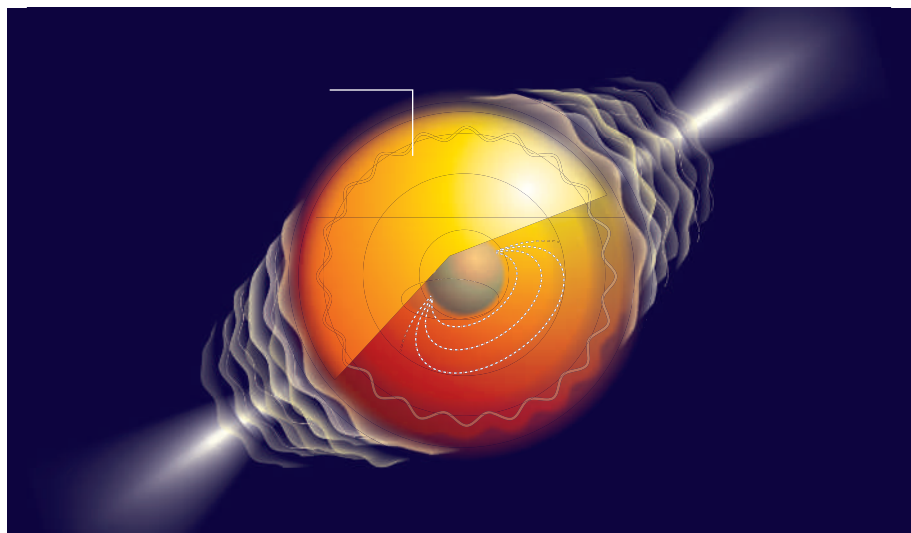


Figure 1 | Shocking past. Data from NASA's Swift telescope, together with ground-based follow-up observations of an object known as SN 2006aj/GRB060218 seem to indicate^{1–4} that γ -ray bursts (GRBs) and supernovae have a common origin in a collapsing Wolf-Rayet star. This object could have a highly magnetized compact object known as a magnetar at its core, which supplies the magnetic energy to produce both a highly relativistic jet (the GRB) and a mildly relativistic expanding shockwave. This shockwave produces radiation at all frequencies, the typical signature of a supernova explosion, when it breaks through the surface of the collapsing star slightly after the faster-moving GRB.

second-closest GRB observed so far — it would have gone unnoticed. Soderberg *et al.*⁴ and Pian *et al.*² therefore both speculate that fainter GRBs, connected with fainter supernovae, occur more frequently than was thought, but are simply being missed.

Further evidence for a hitherto unnoticed GRB-supernova connection comes from the optical luminosity of the event. When, after about two days, the optical afterglow of the GRB had subsided, the overall luminosity began to rise again spectacularly. The extra luminescence is typical of a supernova when radioactive nickel-56 nuclei, synthesized in the initial explosion, decay to cobalt-56. This process reheats the material, causing it to glow at optical frequencies. Such a bump in the light curve is not uncommon in GRBs — in a study of the optical afterglows of 21 GRBs at known distances, nine were found to exhibit a significant excess⁹. Perhaps these were the signatures of accompanying supernovae that were just barely peeking their heads above the GRB afterglow.

Expanding debris from the explosion offers a wealth of information about the mechanism that produced it, and, after three days, ground-based optical telescopes also detected these remnants of the supernova explosion. The lack of hydrogen and helium signatures in the optical spectra classified the supernova as a type Ic, with an ejected mass lower than the norm for such explosions^{2,4,10}.

These observations^{1–4} would thus seem to extend the range of star types that can produce a GRB to lower-mass stars, and might also demand an alternative mechanism capable of producing a GRB jet. Traditionally, accretion on to a black hole is held responsible for this; but lower-initial-mass stars are thought to produce not a black hole, but an extremely compact object of a different flavour — a neutron star.

Soderberg *et al.*⁴ propose a highly magnetized neutron star, known as a magnetar, as the culprit behind the observed explosion (Fig. 1). This would solve the problem^{2–4} that the non-thermal X-ray observations cannot be explained by the usual GRB mechanism. But in fact, the X-ray flux was observed to follow a two-component power-law decay similar to that seen in typical GRB afterglows^{1,10,11}, calling into question the need for a special mechanism. It might be that a magnetar is powering the non-thermal component, but by what means? The lower luminosity of GRB060218 suggests a weak, collimated jet and/or little debris for the jet to interact with. That can be accommodated perfectly well by the traditional jet model of GRB creation, but scaled down to lower energies.

Questions still remain on the detailed geometry and dynamics of the shockwave produced in exploding stars. For GRB060218, it is likely that only the weak nature of the jet revealed the underlying supernova¹². In general, if the relativistic jet dominates, the supernova will be lost in the optical afterglow of the burst. Or

perhaps the star's resources were used up to power the jet, and no supernova occurred. In that case, most of the expelled matter would be seen emanating from the central engine as two cone formations — the classic GRB. If, on the other hand, the (almost) spherical shockwave dominates, a supernova is seen.

What is now clear is that at least some GRBs are shots fired to warn of the imminent explosion of massive stars. They can help us direct our telescopes the right way to learn more about the greatest explosions in the Universe. ■

Timothy R. Young is in the Department of Physics, University of North Dakota, PO Box 7129, Grand Forks, North Dakota 58202, USA.

e-mail: tim_young@und.nodak.edu

1. Campana, S. *et al. Nature* **442**, 1008–1010 (2006).
2. Pian, E. *et al. Nature* **442**, 1011–1013 (2006).
3. Mazzali, P. A. *et al. Nature* **442**, 1018–1020 (2006).
4. Soderberg, A. M. *et al. Nature* **442**, 1014–1017 (2006).
5. Galama, T. J. *et al. Nature* **395**, 670–672 (1998).
6. Hjorth, J. *et al. Nature* **423**, 847–850 (2003).
7. Stanek, K. Z. *et al. Astrophys. J.* **591**, L17–L20 (2003).
8. Malesani, D. *et al. Astrophys. J.* **609**, L5–L8 (2004).
9. Zeh, A., Klose, S. & Hartman, D. H. *Astrophys. J.* **609**, 952–961 (2004).
10. Tagliaferri, G. *et al. Nature* **436**, 985–988 (2005).
11. O'Brien, P. T. *et al.* preprint available at www.arxiv.org/astro-ph/0603530 (2006).
12. Modjaz, M. *et al. Astrophys. J.* **645**, L21–L24 (2006).

NANOSCIENCE

Small talk

Liesbeth Venema

Is rebranding research as 'nanoscience' just jumping on the bandwagon? A recent conference in Basel proves that the name does at least attract researchers from different disciplines to mingle for mutual inspiration.

While most of Switzerland was caught up in the festivities on 1 August that commemorate the birth of the nation, 1,400 scientists gathered in Basel at ICN+T 2006* to discuss the latest developments in nanoscience and technology. They were also there to celebrate a pivotal event in the history of Swiss science: the invention of the scanning tunnelling microscope (STM) at the IBM labs in Zurich more than 25 years ago.

The sharp needle of the STM traces the roughness of surfaces with such high resolution that individual atoms become visible. By opening up a view of the world at nanome-

tre scale, the now-ubiquitous STM catalysed the emergence of a highly diversified field of research — known variously as nanoscience or nano-technology. Exactly where its boundaries lie is notoriously unclear. The researchers who came to Basel fall roughly into two groups: those designing new materials and devices from the bottom up, using nanometre-scale structures as building-blocks; and those creating new tools to investigate matter — mainly biological structures — at the nanometre scale.

Nano is the new micro

Conventional silicon microelectronics will, sooner rather than later, reach a point at which standard structures cannot be scaled down

*International Conference on Nanoscience and Technology, Basel, Switzerland, 31 July–4 August 2006; www.icnt2006.ch

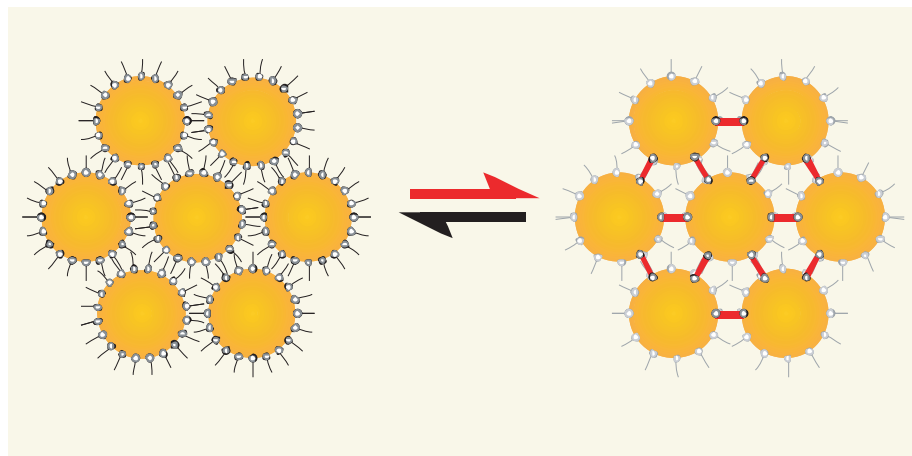


Figure 1 | Forging links. A two-dimensional array of nanoparticles (gold) with insulating spikes becomes electrically linked up after immersion in a solution of conducting organic molecules (red). It is thought that each pair of neighbouring particles is connected by approximately one molecule. The linking process is completely reversible and can be repeated many times. (Figure courtesy of Michel Calame.)

PARASITOLOGY

Peculiar lipid production

African trypanosomiasis — or sleeping sickness — kills about 50,000 people each year according to the World Health Organization. Lee *et al.* (*Cell* **126**, 691–699; 2006) report that the parasite responsible (*Trypanosoma brucei*, pictured) has an unusual way of making a lipid it requires to survive in the blood stream.

The extracellular surface of *T. brucei* is covered in proteins that are attached to the membrane through glycosylphosphatidylinositol (GPI) anchors. The bloodstream form of

T. brucei needs a 14-carbon fatty acid to make the GPI anchors, and this molecule — named myristate — is essential for pathogenesis. Whereas most organisms synthesize fatty acids using type I or type II fatty-acid synthases, Lee *et al.* find that in *T. brucei* myristate is made by a series of enzymes called elongases. As their name implies, these enzymes extend the fatty-acid chain, adding two carbon atoms at a time to a fatty acid that is attached to coenzyme A (CoA).

The authors examined four



candidate elongase (ELO) gene products and determined that ELO1 converts a 4-carbon CoA into a 10-carbon fatty acid; ELO2 uses a 10-carbon CoA to synthesize myristate; ELO3 converts a 14-carbon CoA into an 18-carbon fatty acid; and ELO4 elongates arachidonoyl-CoA into a 22-carbon fatty acid. It was previously known

that ELOs could extend long fatty-acid chains, but this is the first example of a parasite that uses ELOs instead of type I or type II fatty-acid synthases to make a large fraction of its fatty acids.

The authors then examined two related parasites (*Leishmania major* and *Trypanosoma cruzi*) and determined that both organisms contain ELOs that are probably involved in fatty-acid biosynthesis. Additional work is needed to explore how ELOs function *in vivo*, but the authors suggest that it may be possible to exploit this unique pathway to develop new anti-parasitic drugs.

Joshua M. Finkelstein

any further. Researchers hope that by then new-style bottom-up devices will be ripe for commercial use. Designing structures at the nanometre scale allows the quantum-mechanical properties of matter to be exploited and devices with unprecedented functions to be built. Indeed, the ‘qubits’ that are the basic units of quantum computing were recurring stars of the Basel conference — most frequently in their incarnation as single electrons confined within the islands of semiconductor material known as quantum dots. Although it is unlikely that the next generation of computer chips will be based on quantum logic, significant progress is being made in constructing practical qubits to allow logic operations in a small circuit (Lieven Vandersypen, Delft Univ. Technology)¹.

Devices and electronic materials based on the properties of tailor-made molecules are also being designed. In various theoretical proposals and experimental realizations, single molecules sandwiched between two electrical contacts work as miniature electronic devices — as rectifiers, for example, which admit current in only one direction. The leap from constructing a prototype single-molecule device to integrating it into a larger-scale electronic circuit with reliable characteristics is clearly a big one. Here, the principle of ‘self-assembly’ is often exploited, in which anchor groups are chemically attached to molecules, forcing them to grab hold of surfaces and form an organized layer.

An innovative self-assembly approach reported at the conference involved the creation of a two-dimensional network in which organic molecules form electrical links between metal nanoparticles (Laetitia Bernard, Univ. Basel)². First, the nanoparticles order themselves in a neat array, and are electrically isolated from each other by an insulating layer. After immersing this array in a solution of conducting organic molecules, its electrical resistance falls by several orders of magnitude, a drop ascribed to the organic molecules

forming conducting links between the particles (Fig. 1). The organic molecules can be discarded, returning the system to an insulating state, and the high and low values for resistance can be reproduced over many cycles, indicating that a highly ordered network of molecular junctions is formed each time. The researchers claim that their network is a robust platform that can be used to experiment with other tailor-made molecules and nanoparticles and so create more complex nanoscale electronic circuits.

Taking the bioroute

Molecular electronics is an area where physicists have forged fruitful collaborations with chemists. Other physicists have crossed over to the life sciences, where there are many challenging problems to tackle at the single-molecule scale. Among the topics discussed at the conference were sensitive detectors for biological molecules in solution (Scott Manalis, MIT); what kind of tool can unfold single strands of RNA (Cees Dekker, Delft Univ. Technology); and whether a microscope can be constructed to determine the chemical structure of a complex three-dimensional molecule (Dan Rugar, IBM Almaden Research Center).

A basic tool now widely used for imaging biological structures is a cousin of the STM called the atomic force microscope (AFM). Instead of a sharp metallic needle, the AFM uses a thin vibrating strip of material — a cantilever — to feel the contours of a surface, potentially at nanometre resolution. The imaging process is based on the sensitive dependence of the cantilever’s resonance frequency on interaction forces with the surface.

In a variation on this principle, microscale or nanoscale resonators have been found useful as stand-alone devices for sensing molecules or small particles and detecting their mass. At the ultimate detection limit, a nanoscale resonating strip cooled to cryogenic temperatures can weigh clusters of atoms down to a

mass resolution of 7 zeptograms (7×10^{-21} g) (Michael Roukes, Caltech)³. That is equivalent to just 30 atoms of xenon. At the other end of the spectrum, micro- and nanoelectromechanical sensing systems that are cheap, portable, easy to use and have an integrated read-out are being developed for large-scale scientific or commercial applications, such as detecting chemical substances or monitoring biochemical reactions (Anja Boisen, Tech. Univ. Denmark)⁴.

Creative buzz

Many of the early STM enthusiasts were present in Basel, including Heinrich Rohrer and Gerd Binnig, the IBM researchers who shared the 1986 Nobel Prize in Physics for the technique’s invention. Among the anecdotes was the story that the first seminal paper reporting the new technique was rejected by a leading physics journal: hindsight is indeed a wonderful thing. Nowadays, STM and AFM are used in a huge variety of studies, including the effect of green tea on health by imaging cell structure, detecting acoustic bursts in butterfly cocoons and even as one source of ideas for a new form of science-inspired dance, as entertainingly demonstrated in a plenary talk (Jim Gimzewski, UCLA).

Indeed, a remarkable, almost contagious, creative activity in nanoscience was evident at ICN+T 2006. A point emphasized time and again was the importance of doing things that seem crazy at first glance. As one contributor put it at the end of his presentation, “I wonder if this will work, but I am going to try it anyway.”

Liesbeth Venema is a senior editor for the physical sciences at *Nature*.

1. Koppens, F. H. L. *et al.* *Nature* **442**, 766–771 (2006).
2. Liao, J., Bernard, L., Langer, M., Schönenberger, C. & Calame, M. *Adv. Mater.* (in the press).
3. Yang, Y. T., Callegari, C., Feng, X. L., Ekinci, K. L. & Roukes, M. *Nano Lett.* **6**, 583–586 (2006).
4. Dohn, S., Hansen, O. & Boisen, A. *Appl. Phys. Lett.* **88**, 264104 (2006).

OBITUARY

Setsuro Ebashi (1922–2006)

Physiologist who uncovered the regulatory role of calcium in cells.

All animal movement depends on muscle contractions. Nerve impulses cause a change of electrical potential across the muscle-cell membrane, which rapidly propagates to both ends of the cell, causing contraction. How electrical excitation at the muscle cell's surface induces contraction of the proteins packed inside was one of the biggest mysteries in physiology for decades. It was Setsuro Ebashi, who died on 17 July 2006 at the age of 83, who essentially solved this conundrum.

Muscle cells contain two types of filament that are aligned along the long axis of the cell. Contraction occurs when the protein myosin in the 'thick' filaments reacts with another protein, actin, in the 'thin' filaments, causing the filaments to slide across each other, pulling the ends of the cell closer together. The process is fuelled by the breakdown of ATP. The biochemical basis of contraction was largely determined by Albert Szent-Györgyi in the 1940s, and the biophysical and structural basis was uncovered by Jean Hanson and Hugh Huxley, and by Andrew Huxley and R. Niedergerke, during the 1950s. But how the process was triggered by nerves remained a mystery.

Ebashi discovered that in the absence of calcium ions, no contractile reaction occurs, even when ATP is added to the myosin–actin system, but that with even a minute amount of calcium (of the order of 1 micromolar), ATP induces a vigorous contractile reaction. This calcium-dependence had been previously overlooked by biochemists because of low-level contamination from laboratory glassware and impurities in the chemical reagents. Ebashi took tremendous pains to avoid contamination by calcium ions in all the solutions and protein preparations he used, making his results unequivocal. His work was consistent with contemporary results from Annemarie Weber, who showed independently that the breakdown of ATP during the contractile reaction requires minute amounts of calcium.

The idea that calcium might be involved in muscle contraction came to Ebashi during his studies on the 'relaxing factor' reported by B. B. Marsh in 1951. This was a fraction of the homogenate that is made when muscle cells are ground up, and it could induce relaxation of the myosin–actin system. Ebashi proved that relaxing factor is nothing but fragmented pieces of a muscle-specific organelle called the sarcoplasmic reticulum. Enquiring into the factor's mechanism of action, Ebashi got a hint from Emil Bozler's result in 1954 that EDTA — a 'chelating' agent that sequesters

calcium — causes relaxation. Ebashi compared the calcium-binding activity of various chelating agents with their relaxing activity and found that they correlated exactly. He further showed that fragmented sarcoplasmic reticulum can accumulate calcium ions rapidly in the presence of ATP, and so can remove enough calcium from the surrounding medium to cause relaxation.

The relaxation process is the reverse of contraction, so Ebashi proposed what is our current understanding of excitation–contraction coupling: excitation at the surface membrane somehow sends a signal to the sarcoplasmic reticulum; this releases the calcium ions that accumulate there during the relaxing and resting period, and the flood of calcium ions induces the contractile reaction.

Examining the process more closely, Ebashi discovered that purified myosin and actin react with ATP even in the complete absence of calcium ions, and that the regulatory action of calcium is exerted only in the presence of a certain protein factor. This factor turned out to be a mixture of tropomyosin, a protein that previously had no known function, and a newly discovered protein that Ebashi named troponin. Tropomyosin and troponin are present with actin in the thin filaments. In the absence of calcium ions, the two proteins cooperate to inhibit actin, preventing it from interacting with myosin in the thick filaments. Having found that calcium binds strongly to troponin, Ebashi proposed that the resulting conformational changes in troponin are transmitted through tropomyosin to actin to remove the inhibition, and the contractile reaction ensues — a mechanism that has subsequently been confirmed.

In the early 1960s, the idea that a simple inorganic ion such as calcium controls contraction was not popular among most biochemists — the prevailing belief was that such an important biological phenomenon as contraction should be regulated by sophisticated organic molecules. So Ebashi had a hard time getting his ideas taken seriously, despite his clear evidence. Only after the discovery of troponin and his elucidation of the mechanism did everybody accept the regulatory role of calcium ions.

The regulatory roles of calcium are not confined to muscle contraction. Since Ebashi's discovery, numerous cellular processes, including the release of neurotransmitters and hormones, metabolic switching and gene expression, have been found to be controlled by calcium. Thus,



D. E. PHILPOTT

Ebashi really broke open the field of calcium signalling, and profoundly influenced life science as a whole.

Setsuro Ebashi was brilliant from childhood. He skipped a year in both primary and middle schools and was admitted into the Dai-ichi High School, the most prestigious high school in Japan at the time. He was promoted to be professor of pharmacology at the University of Tokyo at the age of 36. But his scientific success was not only due to his sharp mind — he worked extremely hard in the laboratory, until after midnight most nights. Although he had many collaborators, most notably his wife Fumiko Ebashi, all the major results were, amazingly, produced by Ebashi himself.

As one of the most respected scientists in physiology and pharmacology worldwide, Ebashi received numerous honours, including the Order of Cultural Merit, the highest scientific honour in Japan. He was a Member of the Japan Academy and a Foreign Member of the Royal Society in London, and of several other academies, including those of the United States, Germany and Belgium. He also contributed to the global scientific community as president of the International Union for Pure and Applied Biophysics (1978–81) and the International Union of Pharmacology (1990–94), and presided over the International Congress of Pharmacology held in Tokyo in 1981.

Ebashi was a man of attractive personality: charismatic, warm-hearted, helpful and loyal to his colleagues, and patriotic to his country. Despite suffering physical handicaps following a stroke in 2000, he kept a clear mind, and because his condition was fairly stable, nobody expected his sudden death. All his friends and pupils greatly lament it. ■

Makoto Endo

Makoto Endo is in the Department of Pharmacology, University of Tokyo, 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan. e-mail: makoendo@gakushikai.jp

Mast cells are essential intermediaries in regulatory T-cell tolerance

Li-Fan Lu¹, Evan F. Lind¹, David C. Gondek¹, Kathy A. Bennett¹, Michael W. Gleeson¹, Karina Pino-Lagos¹, Zachary A. Scott¹, Anthony J. Coyle², Jennifer L. Reed², Jacques Van Snick³, Terry B. Strom⁴, Xin Xiao Zheng⁴ & Randolph J. Noelle¹

Contrary to the proinflammatory role of mast cells in allergic disorders, the results obtained in this study establish that mast cells are essential in CD4⁺CD25⁺Foxp3⁺ regulatory T (T_{Reg})-cell-dependent peripheral tolerance. Here we confirm that tolerant allografts, which are sustained owing to the immunosuppressive effects of T_{Reg} cells, acquire a unique genetic signature dominated by the expression of mast-cell-gene products. We also show that mast cells are crucial for allograft tolerance, through the inability to induce tolerance in mast-cell-deficient mice. High levels of interleukin (IL)-9—a mast cell growth and activation factor—are produced by activated T_{Reg} cells, and IL-9 production seems important in mast cell recruitment to, and activation in, tolerant tissue. Our data indicate that IL-9 represents the functional link through which activated T_{Reg} cells recruit and activate mast cells to mediate regional immune suppression, because neutralization of IL-9 greatly accelerates allograft rejection in tolerant mice. Finally, immunohistochemical analysis clearly demonstrates the existence of this novel T_{Reg}-IL-9-mast cell relationship within tolerant allografts.

The establishment of immunological tolerance requires both the induction of clonal deletion/anergy and active immune suppression. Immune suppression has been shown to be mediated by unique subsets of T cells called regulatory T cells (T_{Reg})¹. In general, there are two classes of regulatory T cells that impact on peripheral immunity. Naturally arising CD4⁺CD25⁺Foxp3⁺ T_{Reg} cells (nT_{Reg}) are derived from the thymus and have been extensively studied for their roles in autoimmunity and tolerance¹. In contrast, another population of T_{Reg} cells develops from the activation and differentiation of mature CD4⁺CD25⁺ T cells in the periphery; these T_{Reg} cells are called adaptive or induced T_{Reg} cells (iT_{Reg}). In multiple systems of allograft tolerance, the importance of these T_{Reg} cell populations to long-lived allograft survival has been shown by the fact that depletion of T_{Reg} cells before, and even after, the induction of transplantation tolerance results in rapid graft rejection^{2,3}. However, despite the increasing body of knowledge about T_{Reg} cells, the mechanisms by which these cells mediate immune suppression and prevent graft rejection are not well resolved.

Mast cells in tolerant allografts

Mast cells are derived from haematopoietic stem cells, which migrate into vascularized tissues and serosal cavities where they complete their maturation⁴. They are best known as primary responders in allergic reactions such as anaphylaxis and asthma. However, recent studies have shown that mast cells are heterogeneous, can produce an array of both pro- and anti-inflammatory mediators, can act as antigen-presenting cells and express a spectrum of co-stimulatory molecules. These findings indicate that mast cells are far more functionally diverse than previously imagined and can function as immunoregulatory cells that influence both innate and adaptive immunity^{4–6}. Previously, extensive serial analysis of gene expression (SAGE) in tolerant tissue showed that genes predominantly

expressed by mast cells were overexpressed in cultures of activated T_{Reg} cells and in tolerant allografts⁷. These unexpected findings linking mast cells to tolerance prompted our investigation of the potential functional role of mast cells in the establishment of T_{Reg}-cell-mediated allograft tolerance.

To confirm earlier results, the presence of mast cells and their gene products in tolerant allografts in a skin transplantation model was examined. In summary, mice were rendered tolerant to alloantigens by the intravenous infusion of allogeneic cells (a process known as donor-specific transfusion; DST) and concomitant administration of an antibody to CD154 (anti-CD154; ref. 2). This approach allowed for the long-term acceptance of allogeneic skin grafts compared with the non-tolerant control group, which rejected grafts approximately two weeks after grafting. To compare mast-cell-associated gene expression during graft rejection or tolerance, mice received a second graft (30 days after the first grafting), which was harvested seven days later⁷ (Fig. 1a). We then performed quantitative analysis of messenger RNAs in the infiltrating cells extracted from skin transplants. As expected, *Foxp3* and *Il10* expression was highly upregulated in the tolerant group. This indicated the presence of T_{Reg} cells that were producing immunosuppressive mediators in tolerant tissue (Fig. 1b)^{8,9}. In contrast, granzyme B and perforin expression in the rejecting group was much higher than in either the syngeneic or tolerant groups, as shown in other allograft models (Supplementary Fig. 1)¹⁰. Notably, in agreement with previous results⁷, all mast-cell-associated genes examined, including mast cell protease 1 (*Mcpt1*), mast cell protease 5 (*Mcpt5*), tryptophan hydroxylase (*Tph1*) and the high-affinity IgE receptor (*Fcer1a*), were upregulated in the tolerant group compared with the syngeneic and rejecting groups (Fig. 1c).

The mast cell density in syngeneic, rejecting and tolerant grafts was quantified to determine if the increase in mast-cell-gene expression

¹Department of Microbiology & Immunology, Dartmouth Medical School and the Norris Cotton Cancer Center, 1 Medical Center Drive, Lebanon, New Hampshire 03756, USA.

²Department of Autoimmunity and Inflammation, MedImmune, Gaithersburg, Maryland 20878, USA. ³Ludwig Institute for Cancer Research (Brussels Branch) and Experimental Medicine Unit, Université de Louvain, Brussels Branch, 74 Avenue Hippocrate, UCL 7459, B-1200 Brussels, Belgium. ⁴Department of Medicine, Division of Immunology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts 02215, USA.

was due to increased mast cell infiltration in tolerant grafts. Immunohistochemical analysis of rejecting primary grafts at day 7 revealed an apparent increase in infiltrating CD4⁺ T cells relative to that seen in syngeneic or tolerant grafts. In addition, CD117⁺ mast cells that were present in syngeneic grafts were noticeably absent from rejecting grafts at day 7 (Fig. 2a; Table 1). Previously, it was shown that skin mast cells can migrate to the regional lymph node following antigen challenge to the skin¹¹. Therefore, the lack of mast cells in the rejecting allografts could be the result of mast cell migration. However, there was no significant increase in mast cell numbers in the draining lymph nodes in the rejecting group when compared to the syngeneic or tolerant groups (Supplementary Table 1). Hence, the loss of mast cells in rejecting grafts was unlikely to be due to the inflammation-induced migration of mast cells to the regional lymph node, but rather the direct cytotoxic elimination of mast cells by the host. In contrast to rejecting grafts, in established tolerant grafts (day 60) there was massive T_{Reg} (CD4⁺Foxp3⁺)-cell infiltration⁸ (Supplementary Table 2) as well as sustained, or increased, mast cell (CD117⁺) density (Fig. 2b; Table 1). The increased presence of mast

cells in tolerant versus rejecting grafts was consistent with the previous SAGE result, which was indicative of increased mast-cell-gene expression, and the reverse transcription–polymerase chain reaction (RT–PCR) data reported herein. Collectively, these data indicate that both mast cells and T_{Reg} cells increase in number in tolerant allografts and may be crucial in sustaining allograft survival.

No long-term allograft tolerance in mast-cell-deficient mice

Mast cells were functionally implicated in T_{Reg}-cell-mediated allograft survival through a series of studies in mast-cell-deficient mice (C57BL/6 *Kit*^{W^{sh}; *Kit*^{W^{sh} (*W^{sh}*) mice). *W^{sh}* mice have an inversion mutation in the transcriptional regulatory elements upstream of the *c-Kit* open reading frame, which influences *c-Kit* gene expression in a tissue- and age-specific manner^{4,12}. Mast cell numbers in *W^{sh}* mice decrease exponentially after birth owing to their developmental dependence on *c-Kit*^{13,14}, whereas the frequency of other haematopoietic and lymphoid cell populations remains relatively normal. To assess the role of mast cells in allograft survival, *W^{sh}* mice were administered anti-CD154 and DST to induce allospecific tolerance (Fig. 3a) and grafted with allogeneic skin. Co-administration of anti-CD154 and DST induced long-term acceptance of skin allografts in wild-type mice (median survival time (MST) >70 days), but not in the *W^{sh}* mice (MST = 17 days; *P* < 0.0001; log-rank test; Fig. 3b). In fact, anti-CD154/DST-treated *W^{sh}* mice rejected allografts at a rapid pace that was almost indistinguishable from the untreated control mice (MST = 13 days). Hence, mast cells were essential for anti-CD154/DST-induced tolerance. One of the control groups (syngeneic C57BL/6 skin transplant onto *W^{sh}* mice) demonstrated delayed rejection (MST = 45 days) with long-term survival of 40% of the grafts (Fig. 2b). This delayed rejection could be due to differences in minor histocompatibility molecules in addition to the impaired development of melanocytes in the *W^{sh}* mice¹³. *W^{sh}* mice express *c-Kit* in melanocytes at an early age, which should allow the central tolerance to become established; however, the lack of normal melanocyte development in the adult may allow for this minor, delayed rejection response.}}

To examine further the role of mast cells in transplantation tolerance, mast-cell-knockin mice were prepared⁴. Previously, it has been shown that mast cells can be reconstituted both systemically and/or regionally in mast-cell-deficient mice by adoptive transfer of bone-marrow-derived mast cells (BMMCs) generated *in vitro*^{4,14}. BMMCs were harvested five weeks after the initiation of culture (Supplementary Fig. 2a), and a total of 5×10^6 cells (>99% c-Kit⁺FcεRIα⁺) were injected intradermally into the back skin of *W^{sh}* mice. Eight weeks after intradermal reconstitution, mice showed comparable mast cell frequency in back skin to that observed in wild-type C57BL/6 mice, as reported previously¹⁴ (Supplementary Fig. 2b, c). BMMC-reconstituted *W^{sh}* mice were then grafted with skin transplants after anti-CD154/DST co-administration and the acceptance of allografts was monitored over time (Fig. 3c). As shown in Fig. 3d, local reconstitution of mast cells in the back skin was able to extend graft survival on *W^{sh}* mice (MST = 44 days) that were initially unable to sustain allografts following anti-CD154 and DST treatment (MST = 20 days; *P* = 0.0052). Although the rapid

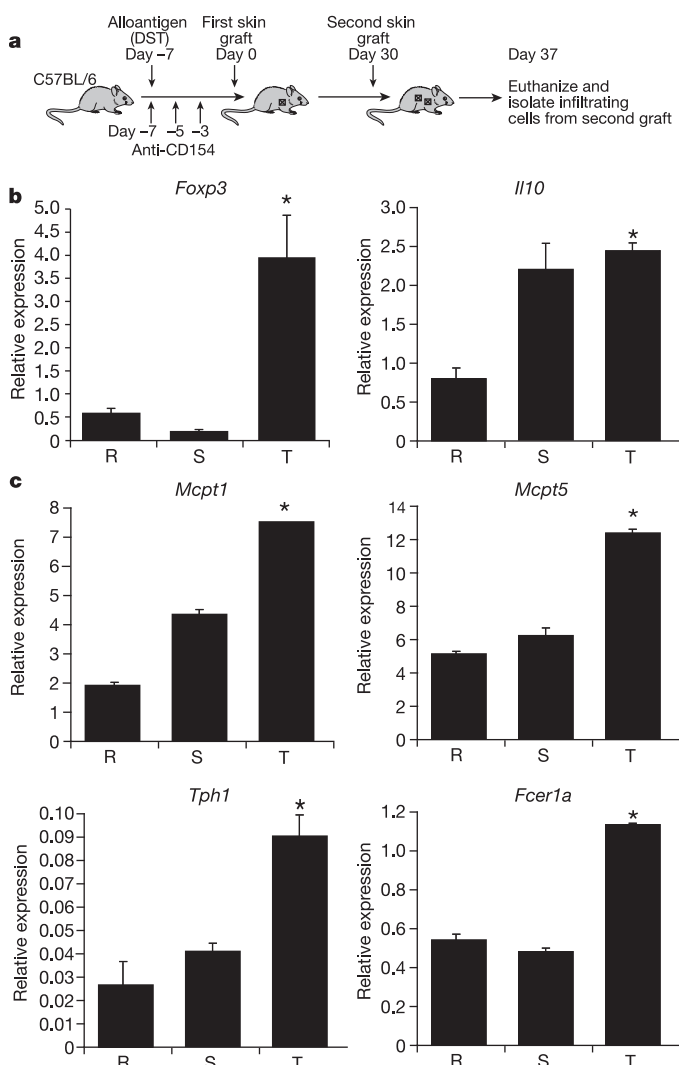


Figure 1 | Mast-cell-related gene expression in tolerant allografts.

a, Allograft tolerance was induced by anti-CD154 and DST. Thirty days after the first skin graft, mice received a second graft, which was harvested 7 days later. **b**, Infiltrating cells from rejecting (R), syngeneic (S) and tolerant (T) grafts were isolated and quantitative real-time RT–PCR (qRT–PCR) analysis was performed. **c**, qRT–PCR analysis of mast-cell-related gene mRNA levels in the indicated groups. Data in **b** and **c** are representative of two individual experiments; values represent the mean $\pm$ s.d., **P* < 0.05 (R versus T).

Table 1 | Numbers of mast cells in skin post-grafting

Day	Mast cells per mm ²		
	Rejecting	Syngeneic	Tolerant
7	19.3 $\pm$ 18.9	85.8 $\pm$ 16.4*	49.0 $\pm$ 24.4*†
60	NA	77.8 $\pm$ 16.8	90.9 $\pm$ 38.8‡§

Data are representative of two individual experiments (*n* = 3–6 mice per group); values represent the mean $\pm$ s.d. NA, not available.

**P* < 0.05 by analysis of variance versus values for R at day 7.

†*P* < 0.05 by analysis of variance versus values for S at day 7.

‡*P* = 0.5005 by analysis of variance versus values for S at day 60.

§*P* < 0.05 by analysis of variance versus values for T at day 7.

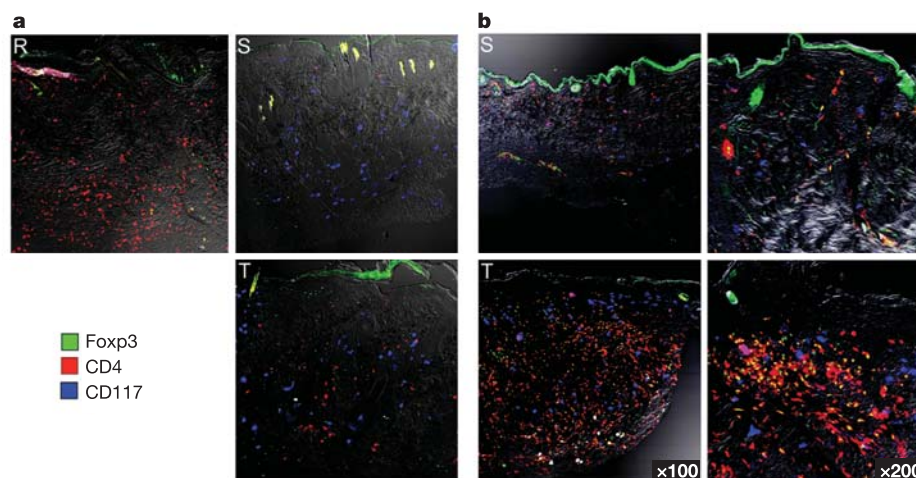


Figure 2 | Histological analysis of cell infiltration in tolerant allografts. **a**, Cryosections of skin transplants from the indicated groups (R, rejecting; S, syngeneic; T, tolerant) 7 days post-grafting were stained for infiltration of CD4⁺ T cells (red), CD4⁺Foxp3⁺ regulatory T cells (yellow) and CD117⁺

skin mast cells (blue). Magnification is $\times 100$. **b**, For transplants at 60 days, a higher original magnification ($\times 200$; right column) was used to show the colocalization of different cell populations.

allograft rejection in the anti-CD154/DST-treated W^{sh} mice indicates a role for mast cells in allograft survival, the data can merely implicate c-Kit⁺ cells. However, together with the mast-cell-reconstitution experiments, these results strongly suggest an indispensable role for mast cells in the establishment of skin transplant tolerance.

IL-9 links mast cells to T_{Reg}-cell-mediated allograft tolerance

The data presented thus far indicated that both T_{Reg} and mast cells were important in long-lived allograft tolerance. However, it remained unclear whether these two cell types functionally interacted to establish regional tolerance in the tolerant allograft. Analysis of gene array data from anti-CD3/anti-glucocorticoid-induced tumour necrosis factor related gene (GITR)-activated T_{Reg} cells indicated that *Il9* was highly upregulated on T_{Reg}-cell activation (Fig. 4a). IL-9 was initially cloned as a T-cell growth factor whose receptor shares the common γ -chain with IL-2 family members such as IL-2, -4, -7, -15

and -21 (refs 15, 16). Subsequently, IL-9 was shown to be a mast cell growth factor on the basis of its capacity to enhance the survival of primary mast cells and to induce their production of inflammatory cytokines, mast cell proteases and the high-affinity IgE receptor (Fc ϵ RI α)¹⁷. Furthermore, IL-9-deficient mice contained far fewer mast cells than their wild-type littermates¹⁸. As shown in Fig. 4a, following anti-CD3/anti-GITR stimulation, which has been shown to have an important role in controlling T_{Reg} cell activities^{19,20}, *Il9* was markedly upregulated in T_{Reg} cells, but not the CD4⁺CD25⁻ T-cell population. Upregulation of IL-9 protein expression in T_{Reg} cells was confirmed by an IL-9 enzyme-linked immunosorbent assay (ELISA; Fig. 4b). Moreover, CD28 co-signalling could induce levels of IL-9 production that greatly exceeded those observed with GITR co-signalling (Fig. 4b). IL-9 production seems to be a common feature of T_{Reg} cells because both nT_{Reg} cells (CD4⁺CD25⁺) as well as iT_{Reg} cells (CD4⁺CD25⁻ T cells cultured with anti-CD3 and transforming

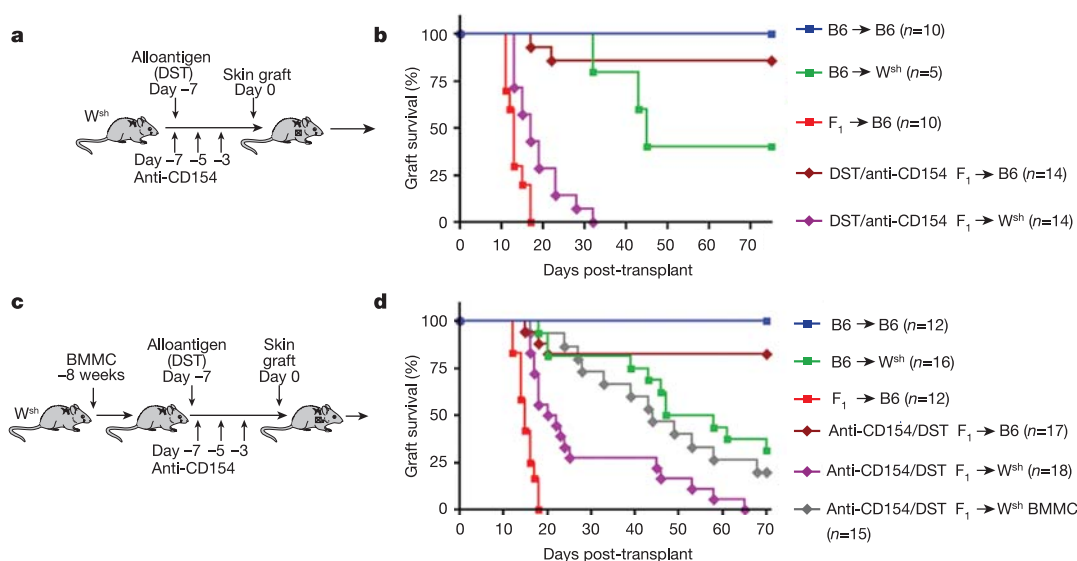


Figure 3 | Mast-cell-deficient mice are not capable of establishing long-term allograft tolerance. **a**, Anti-CD154 and DST were co-administered to mast-cell-deficient mice (W^{sh}) as described in the text. **b**, Allogeneic and syngeneic graft survival in treated W^{sh} and C57BL/6 (B6) mice was followed over time. F₁ is a hybrid of C57BL/6 and BALB/c. **c**, A group of W^{sh} mice

received BMMCs subcutaneously, 8 weeks before anti-CD154/DST co-administration and skin grafting. **d**, Graft survival was followed over time and compared with the control group without BMMC reconstitution. All skin graft rejection assays described above have been confirmed by at least three individual experiments.

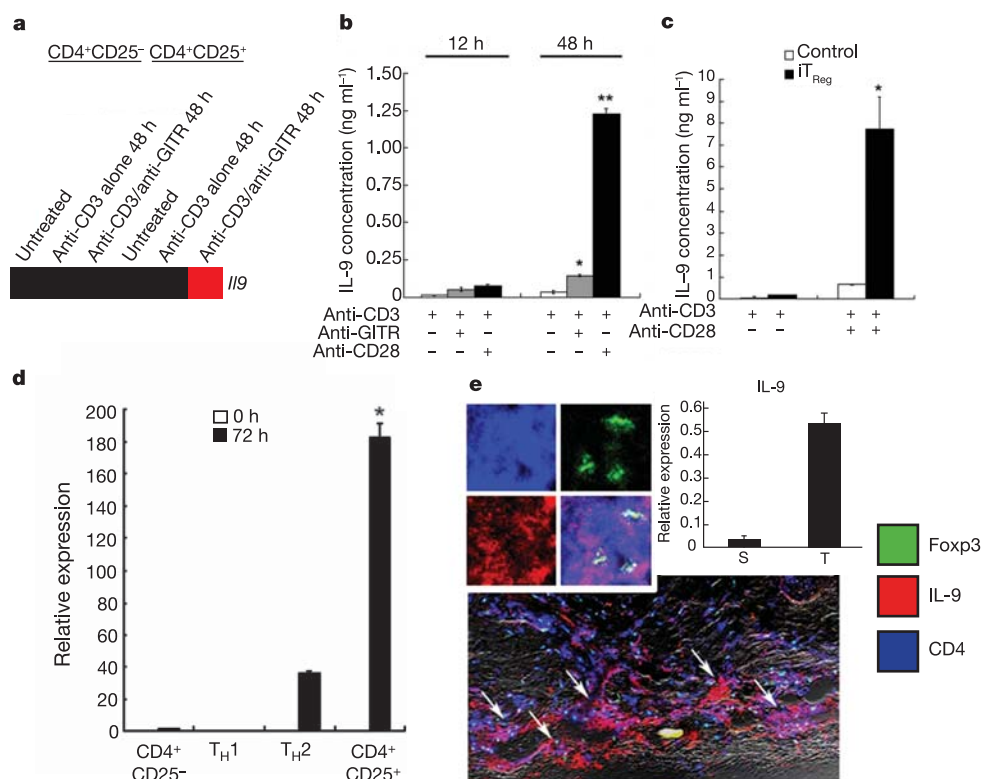


Figure 4 | T_{Reg} cells produce high levels of IL-9 on activation both *in vitro* and *in vivo*. **a**, CD4⁺CD25⁻ and CD4⁺CD25⁺ T cells were cultured *in vitro* with anti-CD3 and IL-2 with or without anti-GITR for 48 h. RNA was hybridized to the Affymetrix A430 array. **b**, IL-9 concentration in T_{Reg} cell culture supernatant with different treatments for 12 h and 48 h. Values represent mean $\pm$ s.d., * P < 0.001 and ** P < 10⁻⁷ versus anti-CD3 alone. **c**, IL-9 production in iT_{Reg} cell culture on activation at 48 h. Values represent mean $\pm$ s.d., * P < 0.05 versus anti-CD3 alone. **d**, qRT-PCR of IL-9 expression in different T-cell subsets. T-cell subsets were stimulated with anti-

CD3 and anti-CD28 for 72 h as described. Values represent mean $\pm$ s.d., * P < 0.001 versus T_H2 cell culture. **e**, Skin transplants from tolerant group 60 days post-grafting were stained for infiltration of CD4⁺ T cells (blue), CD4⁺Foxp3⁺ regulatory T cells (light blue) and IL-9 (red). IL-9-secreting CD4⁺Foxp3⁺ regulatory T cells are shown in purple (white arrows). Magnification is $\times$ 100. Higher magnification (800 $\times$) of IL-9-secreting CD4⁺Foxp3⁺ regulatory T cells are shown with and without merging in the upper left panels. The right panel shows the qRT-PCR for IL-9 mRNA levels of the grafts in the indicated groups (mean $\pm$ s.d.).

growth factor (TGF)- β *in vitro*²¹) produced high levels of IL-9 on activation (Fig. 4c). CD4⁺CD25⁻ T cells cultured in the absence of TGF- β produced <10% of the IL-9 compared with iT_{Reg} cells (Fig. 4c). As reported and shown in Fig. 4d, T-helper-2 (T_H2) cells secreted IL-9 on activation; however, T_{Reg} cells were superior in terms of IL-9 production on a per-cell basis. Notably, although high levels of IL-9 were produced, IL-9 did not seem to have a role in autocrine T_{Reg} cell growth or suppressive activity *in vitro* (Supplementary Fig. 3a, b).

Given the high levels of IL-9 produced by T_{Reg} cells and the role of IL-9 in mast cell homeostasis, a functional link between IL-9, mast cells and allograft tolerance was examined. First, immunohistochemical and quantitative real-time PCR analysis of skin transplants indicated that IL-9 was detected in tolerant grafts but not in syngeneic grafts (Fig. 4e and Supplementary Fig. 4). Second, our functional studies *in vivo* indicated a role for IL-9 in T_{Reg}-cell-mediated suppression. To investigate this, we first applied neutralizing anti-IL-9 to the anti-CD154/DST skin allograft model previously described. However, we could see only a partial effect with regard to alterations in graft rejection kinetics (data not shown). Owing to the complexities of the anti-CD154/DST system, which may obscure the potential involvement of IL-9 in T_{Reg}-cell-mediated tolerance, a Rag^{-/-} reconstitution system was employed that allowed the use of defined, enriched populations of T_{Reg} cells and effector T cells to study allograft survival²². Purified CD8⁺ T cells were transferred with or without purified CD4⁺CD25⁺ T cells into grafted Rag^{-/-} mice (Fig. 5a). As reported, the co-transfer of T_{Reg} cells delayed the onset of graft rejection (MST = 42 days) mediated by CD8⁺ T cells in this model (MST = 19.5 days; P = 0.0243). The T_{Reg}-cell-mediated delay

in graft rejection could be completely reversed with anti-IL-9 treatment (MST = 23 days; P = 0.0161; Fig. 5b). As CD8⁺ T cells did not produce IL-9, and there were no other CD4⁺ T cells in this system, this approach allowed us to speculate that IL-9 production by T_{Reg} cells delayed allograft rejection. To ascertain whether anti-IL-9 administration resulted in a reduction of mast cell accumulation to the tolerant grafts, mast cell numbers in various tissues were quantified. A reduced number of mast cells in the skin of mice treated with anti-IL-9 compared with those treated with control-mouse immunoglobulin was observed at day 10 after the transplantation of the allograft into Rag^{-/-} mice (Supplementary Fig. 5a, b). Hence, the regional production of IL-9 within the tolerant allograft facilitated mast cell accumulation. Although at this time we cannot definitively say IL-9 production by T_{Reg} cells mediates mast cell recruitment and function, IL-9 has clearly been shown to be instrumental in peripheral suppression of alloreactive CD8⁺ T cells.

Discussion

There are a number of novel, unanticipated findings that emerge from the studies presented here. First, even though recent studies have underscored the plasticity of mast cells in regulating acquired immune responses^{4,6,23}, the fact that mast cells may be instrumental in orchestrating T_{Reg}-cell-mediated peripheral tolerance is unprecedented. It is known that host-derived TGF- β is crucial for the peripheral immunosuppression mediated by T_{Reg} cells, and it is tempting to speculate that T_{Reg}-cell-activated mast cells are responsible for TGF- β production, or the liberation and activation of TGF- β via other known or unknown factors that mast cells secrete²⁴. In

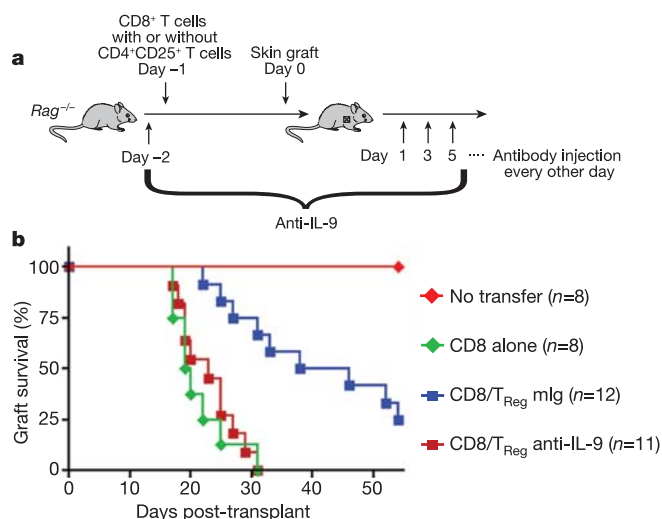


Figure 5 | IL-9 secreted by T_{Reg} cell functionally links mast cells to T_{Reg} -cell-mediated allograft tolerance. **a**, $Rag^{-/-}$ mice were reconstituted with purified $CD8^{+}$ T cells with or without T_{Reg} cells at a 5:1 ratio, 1 day before grafting. In the IL-9-treated group, neutralizing anti-IL-9 monoclonal antibody was administered intraperitoneally every other day starting 2 days before grafting as shown in the figure. **b**, Allograft survival was followed over time and compared with the control group treated with mouse immunoglobulin (mlg). All assays assessing the kinetics of skin graft rejection described above have been confirmed by at least three individual experiments.

addition, TPH1, like indoleamine-pyrrole 2,3-dioxygenase (Indo), is an enzyme that can metabolize tryptophan and create a tryptophan-deficient environment²⁵. As such, this may be a mechanism used by mast cells to limit T-cell activation. Second, recruitment of mast cells into sites of peripheral tolerance may be a common mechanism to control long-lived immune unresponsiveness at that site. In addition to allograft tolerance, a number of tumour models have documented the accumulation of mast cells, as well as T_{Reg} cells, to the tumour sites. Also, it has been shown that mast cells may contribute to tumour growth and metastasis^{23,26}. Thus, like in the allograft model presented, the T_{Reg} cell–mast cell partnership may also have an immunosuppressive role in dampening the immune response to tumours. Third, both n T_{Reg} cells and iT T_{Reg} cells seem to be producers of IL-9, and through the production of IL-9—and other effector molecules—may mediate the activities of mast cells *in vivo*. Hence, IL-9, mast cells and their gene products may become attractive therapeutic targets to ameliorate the impact of T_{Reg} cells *in vivo*.

Active suppression/regulation by T_{Reg} cells is essential to establish and sustain self-tolerance. We present a new paradigm through which T_{Reg} cells operate, and describe a novel set of cells and mediators that control peripheral tolerance. The finding that mast cells are critical in T_{Reg} -cell-dependent allograft tolerance further expands the knowledge of the interplay of different cellular components in controlling immune responses. Future studies will continue to unravel this pathway and will allow us to understand other mediators and cells that ultimately control peripheral immune suppression.

METHODS

Mice. C57BL/6, CB6F₁ (hybrid of C57BL/6 and BALB/c), C57BL/6 *Kit^{W-shi}*; *Kit^{W-shi}* (W^{sh}) and C57BL/6 $Rag^{-/-}$ mice were purchased from the Jackson Laboratory. In all skin transplantation experiments, W^{sh} mice were at least 8 weeks old before grafting, to ensure mast cell deficiency¹³. All animals were maintained in a pathogen-free facility at Dartmouth Medical School.

Skin grafting and immunization. Skin grafting was performed following the procedure described previously². In brief, full-thickness tail skins from CB6F₁ (F_1) donors were transplanted onto the dorsal area of age-matched C57BL/6 recipients. Seven days before skin grafting, 4×10^7 T-cell-depleted splenocytes

from an F_1 donor were transferred into recipients through intravenous injection along with three injections of 250 μ g anti-CD154 monoclonal antibody (clone MR-1) on days -7 , -5 and -3 to induce allograft tolerance. For T_{Reg} cell depletion, 250 μ g of anti-CD25 antibody (clone PC61) was administered through intraperitoneal injection 4 days before skin grafting. For blocking IL-9 activities *in vivo*, 200 μ g of neutralizing anti-IL-9 antibody (clone MM9C1)²⁷ was administered through intraperitoneal injection every other day throughout the duration of the experiments. Control recipients received identical amounts of mouse immunoglobulin.

Skin-infiltrating-cell isolation. Skin infiltrating cells were isolated following the modified protocol described previously⁷. Briefly, secondary-challenge skin transplants from different groups were removed 7 days after grafting. Skin grafts were then cut into small pieces, followed by trypsin digestion at 37 °C for 1 h. The remaining pieces were washed with RPMI 1640 medium over nylon mesh. Cell debris was removed by filtration through a 100- μ m nylon cell strainer and a 40- μ m nylon cell strainer, sequentially. The resulting cell suspension was then washed twice in cold HBSS media and used for further analysis.

Cell preparation, BMMC generation and cell reconstitution. Single-cell spleen suspensions were prepared from 8–10-week-old mice. $CD8^{+}$, $CD4^{+}CD25^{-}$ and $CD4^{+}CD25^{+}$ T cells were purified by magnetic separation with MACS (Miltenyi Biotec) according to the manufacturer's instructions. Enriched cell populations and purified cells were phenotypically analysed by fluorescence-activated cell sorting (FACS). The purity of each population was around 90–95%. To generate different T-cell subsets, purified $CD4^{+}CD25^{-}$ T cells were cultured with plate-bound anti-CD3 monoclonal antibody (clone 145-2C11) at 10 μ g ml⁻¹ and soluble anti-CD28 monoclonal antibody (clone PV-1) at 1 μ g ml⁻¹. For T_H1 cell preparation, recombinant mouse IL-12 (5.0 ng ml⁻¹; PeproTech) with neutralizing anti-IL-4 monoclonal antibody (10 μ g ml⁻¹; clone 11B11; BD Pharmingen) were added; for T_H2 cell preparation, recombinant mouse IL-4 (5.0 ng ml⁻¹; PeproTech) with neutralizing anti-interferon (IFN)- γ monoclonal antibody (10 μ g ml⁻¹; clone 37895.11; R&D Systems) were added. Cells were harvested after 5 days of culture and their purities were verified by real-time PCR analysis of lineage-specific gene expression (*Tbx21* for T_H1 and *Gata3* for T_H2 ; data not shown). For iT T_{Reg} cell preparation, recombinant human TGF- β (1 ng ml⁻¹; PeproTech) and human IL-2 (100 U ml⁻¹; PeproTech) was added. After 5 days of culture, cells were harvested and their purities were verified by FACS analysis of *Foxp3* expression (data not shown). For mast cell reconstitution, BMDCs were generated by culturing bone marrow cells with IL-3 (20 ng ml⁻¹; PeproTech) for 5 weeks as shown previously^{28,29}. The purity was assessed by anti-CD117 (c-Kit) and anti-Fc ϵ RI α staining. A total of 5×10^6 BMDCs were then injected intradermally into the W^{sh} recipients 8 weeks before grafting. For $Rag^{-/-}$ mice reconstitution, 1×10^6 $CD8^{+}$ T cells were adoptively transferred through intravenous injection with or without 2×10^5 $CD4^{+}CD25^{+}$ T cells 1 day before grafting.

Real-time PCR and gene array analysis. Total RNA from isolated skin-infiltrating cells or different T-cell subsets was purified using the RNeasy system (Qiagen). Complementary DNA was then prepared and applied to real-time PCR analysis (SYBR green; BioRad). Relative expression of various gene targets normalized to β -actin was calculated as:

$$(2 - (\text{experimental CT} - \beta\text{-actin CT})) \times 1,000$$

where CT is the cycle threshold of signal detection. For gene array analysis, as shown in the previous literature³⁰, RNAs purified from different cell populations with various treatments were analysed using Affymetrix mouse genome A430 oligonucleotide arrays.

Cytokine secretion assay and immunohistology. Secretion of IL-9 was assayed by ELISA. Different T-cell populations (1×10^6) were cultured in 24-well plates precoated with 1 μ g ml⁻¹ anti-CD3 (clone 2C11) with or without 10 μ g ml⁻¹ anti-GITR (also known as Tnfrsf18; clone DTA-1) or anti-CD28 (clone PV-1). Supernatants were collected at the time indicated. IL-9 was quantified according to the manufacturer's instructions (Peprotech). For immunohistology, previously grafted skins were snap frozen, cryocut, and acetone-fixed. Slides were blocked with normal mouse serum. Tissue sections were stained for CD4 (clone GK1.5), CD117 (clone 2B8) and Foxp3 (clone FJK.16s). For IL-9 staining, biotinylated rabbit polyclonal anti-mouse IL-9 antibodies (Peprotech) and PE-conjugated streptavidin (eBioscience) were used. The specificity of IL-9 staining has been confirmed by the absence of staining in skin tissue from *Il9^{-/-}* mice (our unpublished data).

Received 16 May; accepted 20 June 2006.

Published online 20 August 2006.

1. Sakaguchi, S. Naturally arising Foxp3-expressing $CD25^{+}CD4^{+}$ regulatory T cells in immunological tolerance to self and non-self. *Nature Immunol.* 6, 345–352 (2005).

2. Quezada, S. A. *et al.* Analysis of the underlying cellular mechanisms of anti-CD154-induced graft tolerance: the interplay of clonal anergy and immune regulation. *J. Immunol.* **175**, 771–779 (2005).
3. Taylor, P. A., Noelle, R. J. & Blazar, B. R. CD4⁺CD25⁺ immune regulatory cells are required for induction of tolerance to alloantigen via costimulatory blockade. *J. Exp. Med.* **193**, 1311–1318 (2001).
4. Galli, S. J., Nakae, S. & Tsai, M. Mast cells in the development of adaptive immune responses. *Nature Immunol.* **6**, 135–142 (2005).
5. Galli, S. J. *et al.* Mast cells as “tunable” effector and immunoregulatory cells: recent advances. *Annu. Rev. Immunol.* **23**, 749–786 (2005).
6. Benoist, C. & Mathis, D. Mast cells in autoimmune disease. *Nature* **420**, 875–878 (2002).
7. Zelenika, D. *et al.* The role of CD4⁺ T-cell subsets in determining transplantation rejection or tolerance. *Immunol. Rev.* **182**, 164–179 (2001).
8. Graca, L., Cobbold, S. P. & Waldmann, H. Identification of regulatory T cells in tolerated allografts. *J. Exp. Med.* **195**, 1641–1646 (2002).
9. Hara, M. *et al.* IL-10 is required for regulatory T cells to mediate tolerance to alloantigens *in vivo*. *J. Immunol.* **166**, 3789–3796 (2001).
10. Corti, B. *et al.* Potential of real-time PCR assessment of granzyme B and perforin up-regulation for rejection monitoring in intestinal transplant recipients. *Transplant. Proc.* **37**, 4467–4471 (2005).
11. Wang, H. W., Tedla, N., Lloyd, A. R., Wakefield, D. & McNeil, P. H. Mast cell activation and migration to lymph nodes during induction of an immune response in mice. *J. Clin. Invest.* **102**, 1617–1626 (1998).
12. Berrozpe, G. *et al.* The *W^{sh}*, *W⁵⁷*, and *Ph Kit* expression mutations define tissue-specific control elements located between –23 and –154 kb upstream of *Kit*. *Blood* **94**, 2658–2666 (1999).
13. Yamazaki, M. *et al.* *C-kit* gene is expressed by skin mast cells in embryos but not in puppies of *W^{sh}/W^{sh}* mice: age-dependent abolishment of *c-kit* gene expression. *Blood* **83**, 3509–3516 (1994).
14. Grimbaldston, M. A. *et al.* Mast cell-deficient *W-sash c-kit* mutant *Kit^{W-sh/W-sh}* mice as a model for investigating mast cell biology *in vivo*. *Am. J. Pathol.* **167**, 835–848 (2005).
15. Van Snick, J. *et al.* Cloning and characterization of a cDNA for a new mouse T cell growth factor (P40). *J. Exp. Med.* **169**, 363–368 (1989).
16. Jarnicki, A. G. & Fallon, P. G. T helper type-2 cytokine responses: potential therapeutic targets. *Curr. Opin. Pharmacol.* **3**, 449–455 (2003).
17. Zhou, Y., McLane, M. & Levitt, R. C. Th2 cytokines and asthma. Interleukin-9 as a therapeutic target for asthma. *Respir. Res.* **2**, 80–84 (2001).
18. Townsend, J. M. *et al.* IL-9-deficient mice establish fundamental roles for IL-9 in pulmonary mastocytosis and goblet cell hyperplasia but not T-cell development. *Immunity* **13**, 573–583 (2000).
19. Shimizu, J., Yamazaki, S., Takahashi, T., Ishida, Y. & Sakaguchi, S. Stimulation of CD25⁺CD4⁺ regulatory T cells through GITR breaks immunological self-tolerance. *Nature Immunol.* **3**, 135–142 (2002).
20. Tone, M. *et al.* Mouse glucocorticoid-induced tumor necrosis factor receptor ligand is costimulatory for T cells. *Proc. Natl Acad. Sci. USA* **100**, 15059–15064 (2003).
21. Chen, W. *et al.* Conversion of peripheral CD4⁺CD25[–] naive T cells to CD4⁺CD25⁺ regulatory T cells by TGF- β induction of transcription factor *Foxp3*. *J. Exp. Med.* **198**, 1875–1886 (2003).
22. Jarvinen, L. Z., Blazar, B. R., Adeyi, O. A., Strom, T. B. & Noelle, R. J. CD154 on the surface of CD4⁺CD25⁺ regulatory T cells contributes to skin transplant tolerance. *Transplantation* **76**, 1375–1379 (2003).
23. Theoharides, T. C. & Conti, P. Mast cells: the Jekyll and Hyde of tumor growth. *Trends Immunol.* **25**, 235–241 (2004).
24. Mesples, B., Fontaine, R. H., Lelievre, V., Launay, J. M. & Gressens, P. Neuronal TGF- β 1 mediates IL-9/mast cell interaction and exacerbates excitotoxicity in newborn mice. *Neurobiol. Dis.* **18**, 193–205 (2005).
25. Lee, G. K. & Mellor, A. L. Tryptophan deprivation sensitizes activated T cells to apoptosis prior to cell division. *Immunology* **107**, 452–460 (2002).
26. Wedemeyer, J. & Galli, S. J. Decreased susceptibility of mast cell-deficient *Kit^W/Kit^{W-v}* mice to the development of 1,2-dimethylhydrazine-induced intestinal tumors. *Lab. Invest.* **85**, 388–396 (2005).
27. Khan, W. I. *et al.* Modulation of intestinal muscle contraction by interleukin-9 (IL-9) or IL-9 neutralization: correlation with worm expulsion in murine nematode infections. *Infect. Immun.* **71**, 2430–2438 (2003).
28. Razin, E. Culture of bone marrow-derived mast cells: a model for studying oxidative metabolism of arachidonic acid and synthesis of other molecules derived from membrane phospholipids. *Methods Enzymol.* **187**, 514–520 (1990).
29. Saitoh, S. *et al.* LAT is essential for Fc ϵ RI-mediated mast cell activation. *Immunity* **12**, 525–535 (2000).
30. Gondek, D. C., Lu, L. F., Quezada, S. A., Sakaguchi, S. & Noelle, R. J. Cutting edge: contact-mediated suppression by CD4⁺CD25⁺ regulatory cells involves a granzyme B-dependent, perforin-independent mechanism. *J. Immunol.* **174**, 1783–1786 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Galli for critical review of the manuscript during preparation. This work was supported by NIH grants; J.V.S.'s research was funded in part by the Belgian Federal Service for Scientific, Technical and Cultural Affairs and by the Actions de Recherche Concertées, Communauté française de Belgique, Direction de la recherche scientifique.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.J.N. (rjn@dartmouth.edu).

Karyopherin-mediated import of integral inner nuclear membrane proteins

Megan C. King^{1*}, C. Patrick Lusk^{1*} & Günter Blobel¹

Targeting of newly synthesized integral membrane proteins to the appropriate cellular compartment is specified by discrete sequence elements, many of which have been well characterized. An understanding of the signals required to direct integral membrane proteins to the inner nuclear membrane (INM) remains a notable exception. Here we show that integral INM proteins possess basic sequence motifs that resemble 'classical' nuclear localization signals. These sequences can mediate direct binding to karyopherin- α and are essential for the passage of integral membrane proteins to the INM. Furthermore, karyopherin- α , karyopherin- β 1 and the Ran GTPase cycle are required for INM targeting, underscoring parallels between mechanisms governing the targeting of integral INM proteins and soluble nuclear transport. We also provide evidence that specific nuclear pore complex proteins contribute to this process, suggesting a role for signal-mediated alterations in the nuclear pore complex to allow for passage of INM proteins along the pore membrane.

Integral membrane proteins of the INM contribute to the organization of lamins and chromatin at the nuclear envelope¹; mutations in genes encoding lamins and lamin-associated integral INM proteins are tied to a clinically diverse array of genetic diseases referred to as laminopathies². To reach the INM, integral membrane proteins must move past nuclear pore complexes (NPCs), large protein channels that regulate the nucleocytoplasmic transport of soluble molecules³. Because the outer nuclear membrane (ONM) is contiguous and functionally equivalent with the endoplasmic reticulum (ER), INM proteins might reach the cytoplasmic aspect of NPCs by diffusion. Indeed, it has been proposed that integral INM proteins diffuse laterally along the pore membrane and are then selectively retained in the nucleus through interactions with elements of the nuclear architecture⁴. However, as they pass along the pore membrane both transmembrane segment(s) and hydrophilic domains probably encounter elements of the NPC, which may need to restructure locally to accommodate their passage. A recent study in mammalian cells demonstrates that transport of integral membrane proteins to the INM requires energy⁵, suggesting the presence of an active transport pathway in lieu of, or in addition to, passive diffusion/retention.

Heh1 and Heh2 are yeast INM proteins

In order to examine further the mechanism by which integral membrane proteins are targeted to the INM, we searched for potential INM proteins in the model organism *Saccharomyces cerevisiae*. We identified two yeast INM proteins based on homology with the mammalian INM paralogues MAN1 and LEM2. Common to both the yeast and human proteins are an amino-terminal region likely to form a helix-extension-helix (HEH) fold, two transmembrane domains, and a region of homology after the second transmembrane domain (HEH/MAN1 carboxy-terminal homology domain, CTHD)⁶ (Fig. 1a). To reflect this homology, we now refer to these genes as *HEH1* (formerly called *SRC1* (ref. 7)) and *HEH2* (systematic name YDR458C). Both Heh1 and Heh2 localize exclusively to the nuclear envelope when produced in yeast as yellow

fluorescent protein (YFP) fusions at endogenous (Supplementary Fig. 2a) and elevated levels of expression (Fig. 1b and Supplementary Fig. 2b). This contrasts with the distribution of a typical ER protein, which evenly distributes between the perinuclear ER and the cortical ER (Supplementary Fig. 2c). To ensure that Heh1 and Heh2 accumulate specifically in the INM, we performed immunoelectron microscopy. As shown in Fig. 1c, when immunoelectron microscopy was performed using antibodies directed against the C-terminal YFP moiety of Heh2-YFP, gold particles were found almost exclusively along the inside of the nuclear envelope (see Fig. 1d for higher magnification). We determined that over 50% of Heh1-YFP and Heh2-YFP was localized at the nuclear envelope in wild-type cells, and of that greater than 70% was found at the INM (Supplementary Fig. 2d). Thus, Heh1 and Heh2 are useful examples with which to examine the targeting of integral INM proteins.

By further analysing the amino acid sequence of Heh1 and Heh2 and their human orthologues, MAN1 and LEM2, we uncovered the presence of sequences similar to nuclear localization signals (NLSs, Fig. 1a; sequences in Supplementary Table 1), sequence elements that promote active nuclear import of soluble proteins^{8,9}. This finding raised the possibility that mediators of soluble nuclear transport may contribute to INM targeting.

The Ran cycle is required for INM targeting

We first developed an inducible system for monitoring the targeting of Heh1 and Heh2 as YFP fusions produced under the control of the *GAL1* promoter. This allowed us to monitor conditionally the fate of newly synthesized cargo molecules in live cells. To test the hypothesis that mediators of soluble transport contribute to integral INM protein targeting, we examined Heh1-YFP and Heh2-YFP localization in yeast mutants in which nuclear transport is globally inhibited through the disruption of the GTP loading and hydrolysis of the GTPase Ran. GTP-bound in the nucleus and GDP-bound in the cytoplasm, Ran coordinates the bidirectional transport of macromolecules across the nuclear envelope^{8,10}. Proper cycling of Ran (and thus nuclear transport) is disrupted at 34 °C in yeast strains carrying

¹Laboratory of Cell Biology, Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA.

*These authors contributed equally to this work.

temperature-sensitive mutations in the Ran guanine nucleotide exchange factor (*Mtr1/Srm1/Prp20*) or the Ran GTPase-activating protein (*Rna1*). When *Heh1*-YFP and *Heh2*-YFP were produced in a strain harbouring a mutation in *RanGEF* (*mtr1-1*)¹¹ they localized primarily to the nuclear envelope at the permissive temperature (25 °C). Notably, if *HEH1-YFP* and *HEH2-YFP* induction was preceded by shifting cells to 34 °C, *Heh1*-YFP and *Heh2*-YFP accumulated at the periphery of the cell, probably in the cortical ER (Fig. 1e). Using immunoelectron microscopy, we confirmed that

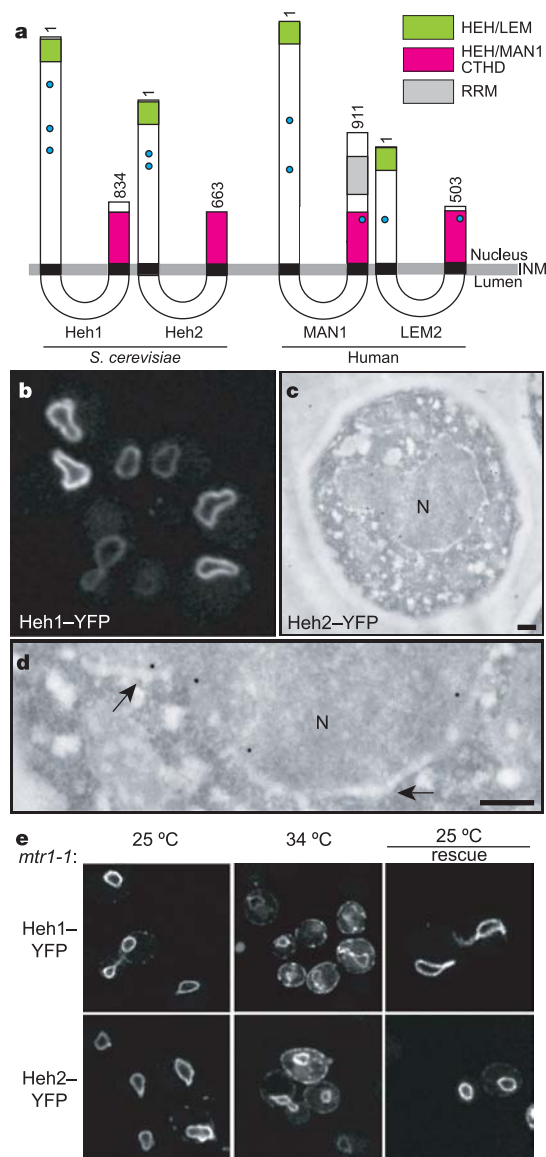


Figure 1 | Heh1-YFP and Heh2-YFP localize to the INM in a Ran-dependent manner. **a**, Diagram of the likely topology of the indicated proteins in the INM. Conserved domains are coloured green (HEH/LEM domain) and magenta (HEH/MAN1 CTHD); transmembrane segments are in black. The RNA recognition motif (RRM) of MAN1 is coloured grey. Cyan circles indicate putative nuclear localization signals. Numbers represent amino acid residues. **b**, Fluorescence micrograph of *Heh1*-YFP in a wild-type yeast strain (CPL160). **c**, Immunoelectron micrograph of *Heh2*-YFP in a wild-type yeast strain (CPL148) labelled with anti-GFP antibodies followed by 10-nm-diameter gold-particle-conjugated secondary antibody. N denotes the nucleus. **d**, Detail of electron micrograph in **c**. Arrows indicate the nuclear envelope. Scale bar in **c**, **d**, 100 nm. **e**, Fluorescence micrographs of *Heh1*-YFP and *Heh2*-YFP produced in the *RanGEF* mutant (*mtr1-1*) strain (CPL162 and CPL163) at the indicated temperature. See Methods for details of rescue.

the cortical ER accumulation correlates with a disruption in INM targeting, as the percentage of *Heh1*-YFP at the nuclear envelope dropped markedly from over 50% in wild-type cells to less than 5% in the *mtr1-1* strain at 34 °C, with essentially no label at the INM (Supplementary Fig. 3a). There was also an appreciable defect in targeting of *Heh1*-YFP in *mtr1-1* cells at the permissive temperature (25 °C, compare Fig. 1b, e and see Supplementary Fig. 3a). Notably, mislocalization of *Heh1*-YFP and *Heh2*-YFP was reversible, as shifting of the culture back to 25 °C (under repressive conditions) led to recovery of nuclear envelope localization (Fig. 1e). Similar mislocalization of *Heh1*-YFP was observed in a strain harbouring a temperature-sensitive mutation in *RanGAP* (*Rna1*), and there were no changes to the distribution of *Heh1*-YFP in wild-type cells at 34 °C (Supplementary Fig. 3b, c).

Nuclear envelope targeting requires karyopherins

Ran regulates nuclear transport by modulating soluble nuclear transport factors called karyopherins^{8,10}. The 'classical' nuclear import pathway is initiated when the cNLS (the 'classical' NLS refers to the NLS in the SV40 T-antigen¹², see Supplementary Table 1) of a cargo molecule is recognized by the import receptor karyopherin- α (importin- α /Kap60/Srp1) in an interaction that is stabilized by binding of karyopherin- β 1 (importin- β 1/Kap95). This ternary complex traverses the NPC and is disassembled by binding to Ran-GTP within the nucleus. To examine whether karyopherins are involved in *Heh1*/*Heh2* targeting, we monitored the localization of *Heh1*-YFP and *Heh2*-YFP in strains with temperature-sensitive mutations or deletions in various karyopherin genes. At the non-permissive temperature in strains harbouring mutations in either *KAP60* (*srp1-31*)¹³ or *KAP95* (*kap95-L63A*)¹⁴, *Heh1*-YFP and *Heh2*-YFP failed to accumulate exclusively at the nuclear periphery, and localized throughout the cortical ER (Fig. 2a and data not shown). We did not detect any defects in the localization of *Heh1*-YFP or *Heh2*-YFP in strains harbouring mutations or deletions in other karyopherin genes including *KAP121* (*kap121-34* (ref. 15), Fig. 2a),

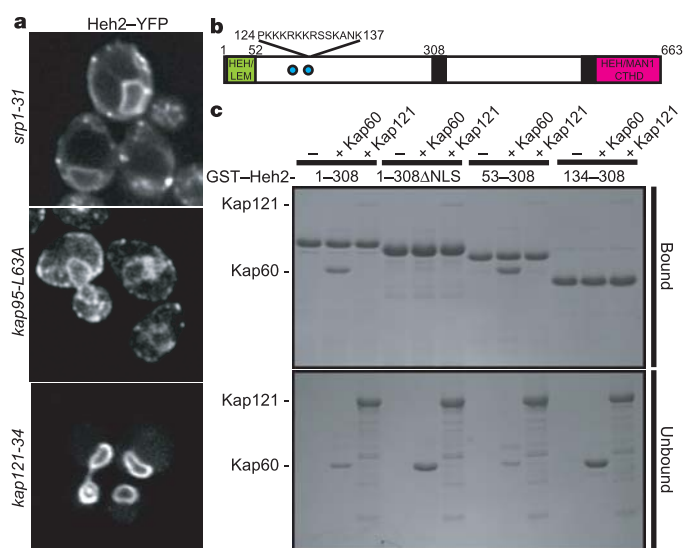


Figure 2 | Kap60 directly interacts with Heh2 and is required for nuclear envelope targeting. **a**, Fluorescence images of the subcellular distribution of *Heh2*-YFP produced in the indicated karyopherin mutants (CPL165, CPL166, CPL167) at the non-permissive temperature (34 °C). **b**, Schematic of *Heh2* with coloration as in Fig. 1a. The amino acid sequence of the *Heh2* NLS is given; numbers are amino acid residues. **c**, The binding of purified recombinant Kap60 and Kap121 to GST fusions of the indicated truncations of *Heh2* was assessed by GST pull-down assays. Equivalent amounts of bound (upper panel) and unbound (lower panel) proteins were separated by SDS-PAGE and visualized by staining with Coomassie blue.

KAP114, KAP120 and KAP122 (data not shown), indicating that Kap60 and Kap95 are specifically involved in targeting of Heh1 and Heh2.

In order to ensure that the mistargeting of Heh1 and Heh2 in the Ran cycle and karyopherin mutants was a direct result of the abrogation of a karyopherin–cargo interaction, we analysed the ability of Heh1 and Heh2 to bind directly to karyopherins. Heh2 has the most canonical cNLS-like sequence¹² in its N-terminal nucleoplasmic domain (Fig. 2b and Supplementary Table 1). Notably, this Heh2 NLS acted autonomously as an NLS *in vivo* as it was able to target GFP to the nucleus in a manner dependent on Kap60 (see Supplementary Fig. 4a). We therefore tested for direct binding between the N-terminal region of Heh2 (Heh2(1–308)) and Kap60 in an *in vitro* GST pull-down assay using recombinant proteins. As shown in Fig. 2c, GST–Heh2(1–308) can bind Kap60, but not Kap121. Although deletion of the HEH domain (Heh2(53–308)) did not affect Kap60 binding, further deletion of residues 1–133 (Heh2(134–308), including all putative NLSs) prevented the interaction with Kap60. Moreover, deletion of the Heh2 NLS alone (Heh2(1–308ΔNLS)) also prevented Kap60 binding.

NLS-mediated targeting of Heh2 is Kap60-dependent

The requirement for Kap60 and Kap95 in Heh2 import *in vivo* and the ability of the Heh2 NLS to mediate binding to Kap60 *in vitro* clearly predict that deletion of the Heh2 NLS should also disrupt INM targeting of Heh2. Consistent with this prediction, a YFP fusion of a mutant of Heh2 lacking the NLS (Heh2(ΔNLS), Fig. 3a), or a point mutant within the Heh2 NLS (Heh2(K126T), Supplementary Fig. 4b and Supplementary Table 1), was mislocalized in wild-type cells and accumulated in the ER. Further analysis by immunoelectron microscopy revealed that Heh2(ΔNLS)–YFP appeared to be excluded from the INM, and was found on membrane stacks contiguous with the ONM, as well as in the cortical ER (Fig. 3b, c). Together, these results indicate that the Heh2 NLS mediates an interaction with Kap60, and that this interaction is required for INM localization of Heh2.

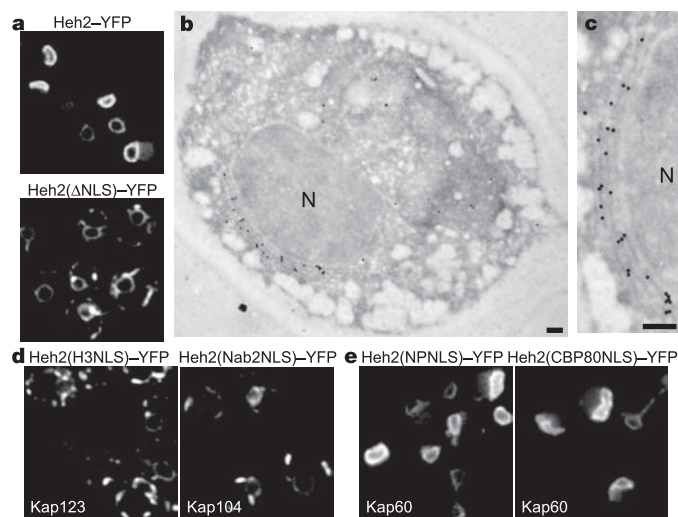


Figure 3 | Kap60 NLSs are required for the INM localization of Heh2.

a, Fluorescence image of Heh2–YFP and a mutant of Heh2 lacking the NLS sequence (Heh2(ΔNLS)–YFP) produced in a wild-type yeast strain (CPL161, CPL176). **b**, Immunoelectron micrograph of Heh2(ΔNLS)–YFP labelled with anti-GFP antibodies followed by 10-nm-diameter gold-particle-conjugated secondary antibody. N denotes the nucleus. **c**, Detail of electron micrograph in **b**. Scale bar in **b**, **c**, 100 nm. **d**, **e**, Fluorescence micrographs of mutants of Heh2 in wild-type yeast strains (CPL100–103) where the Heh2 NLS has been replaced with either Kap123 (Heh2(H3NLS)–YFP), Kap104 (Heh2(Nab2NLS)–YFP) (**d**) or Kap60 (Heh2(NPNLS)–YFP and Heh2(CBP80NLS)–YFP) (**e**) NLSs. NLSs are listed in Supplementary Table 1.

We next investigated whether the NLS-mediated targeting of Heh2 was specific for the Kap60–Kap95 complex, or whether this function could be supported by other members of the karyopherin family. For these experiments, we replaced the Heh2 NLS within the context of full-length Heh2 with NLSs recognized by Kap123, Kap104 or Kap121 (from histone H3, Nab2 and Pho4, respectively; see Supplementary Table 1 for NLS sequences). None of these NLSs could restore the exclusive nuclear envelope localization of Heh2, as these mutants were distributed in a pattern similar to Heh2(ΔNLS)–YFP (Fig. 3d). However, when we replaced the Heh2 NLS with NLSs that use the Kap60–Kap95 pathway, yet clearly differed in their primary sequence, we observed localization of Heh2–YFP at the nuclear envelope (Fig. 3e). Thus, our data support a model whereby the targeting of Heh2 to the INM is a function that it is specific for

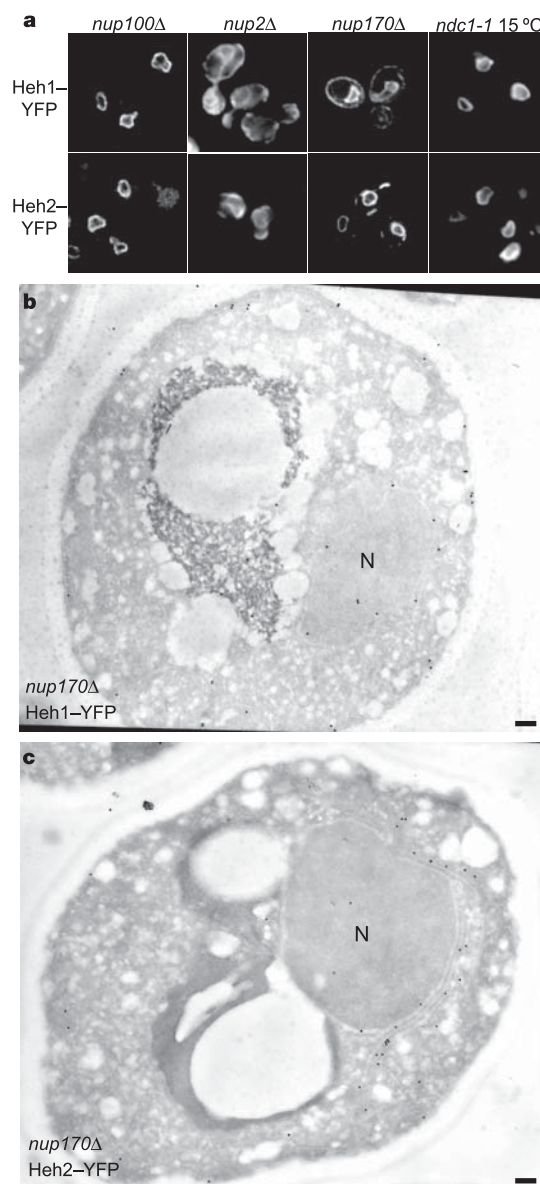


Figure 4 | Heh1–YFP and Heh2–YFP are mistargeted in the absence of Nup2 and Nup170. **a**, Fluorescence micrographs of Heh1–YFP and Heh2–YFP produced in the indicated nucleoporin knockout strains (CPL168–169, CPL177–178, CPL174–175). The cold-sensitive *ndc1-1* mutant strains (CPL179–180) were shifted to 15 °C before galactose induction. **b**, **c**, Immunoelectron micrographs of Heh1–YFP (**b**) or Heh2–YFP (**c**) in a *nup170Δ* strain (CPL174–175) labelled with anti-GFP antibodies followed by 10-nm-diameter gold-conjugated secondary antibody. N denotes the nucleus. Scale bar in **b**, **c**, 100 nm.

Kap60–Kap95 and suggest that this may be a unique function for these karyopherins.

Specific nucleoporins facilitate Heh1/Heh2 import

The involvement of the soluble nuclear transport machinery in the targeting of INM proteins suggests that the hydrophilic domains of INM proteins might be actively transported through the NPC using a mechanism similar to those used by soluble cargoes. Each NPC is composed of ~30 proteins, termed nucleoporins³. We first tested for the involvement of nucleoporins containing phenylalanine-glycine (FG) repeats, which are known to interact with karyopherins¹⁶. Second, because integral membrane proteins must pass along the pore membrane, we tested for the involvement of pore membrane proteins and likely pore-membrane-protein-associated non-FG nucleoporins, such as Nup170 and Nup188 (refs 17–19). Although localization of Heh1–YFP and Heh2–YFP was unaffected by deletion/mutation of most nucleoporins we tested (*nup100Δ*, *ndc1-1* (ref. 20), Fig. 4a; *nup188Δ*, *pom152Δ*, *nup116Δ*, *nup53Δ*, *pom34Δ*, data not shown), we observed aberrant nuclear envelope targeting in strains with gene deletions in *NUP2* and *NUP170*, with Heh1–YFP and Heh2–YFP accumulating in the cortical ER (Fig. 4a). We used immunoelectron microscopy to investigate the distribution of Heh1–YFP and Heh2–YFP in the *nup170Δ* strain in greater detail. In the absence of *NUP170*, Heh1–YFP was almost entirely excluded from the INM (less than 5% was observed at the INM), and was distributed on the ONM and cortical ER (Fig. 4b). Similarly, Heh2–YFP accumulated in the cortical ER and on ONM stacks similar to the pattern observed in cells expressing Heh2(ΔNLS)–YFP (compare Figs 3b and 4c).

Discussion

Here we present evidence that a cNLS can constitute the signal responsible for targeting of integral membrane proteins to the INM. As is the case for soluble cargo, these cNLSs promote import of integral membrane protein cargo in a Ran-cycle- and karyopherin-dependent manner. Although it remains possible that retention signals contribute to the localization of certain integral INM proteins^{21–23}, the data presented here clearly support an active, NLS-mediated pathway. Our data also suggest that the energy requirements in INM targeting⁵ may be contributed by the Ran cycle. NLS-like sequences are found in the majority of mammalian INM proteins including lamin B receptor, LAP1, LAP2β, emerin, MAN1 and LEM2 (ref. 24), in addition to three other potential yeast INM proteins: Prm3, Nem1 and Ydl089w. Trafficking of Prm3 to the nuclear envelope also requires a functional Ran cycle²⁵ (Supplementary Fig. 5a), and, like Heh2, exclusive nuclear envelope localization of Ydl089w at endogenous levels was disrupted by deletion of its putative NLS (Supplementary Fig. 5b and Supplementary Table 1).

Our data suggest that only NLSs recognized by Kap60 (the sole karyopherin-α homologue in yeast) are capable of targeting integral membrane proteins to the INM. The involvement of Nup2 is consistent with this idea, as there are numerous functional links between karyopherin-α and Nup2 (refs 26–29). Intriguingly, a recent report documents the interaction between a karyopherin-α homologue and a viral INM protein as it moves through the ER translocon, suggesting that similar specificity may exist in higher eukaryotes³⁰. This specificity hints at the presence of a unique karyopherin-α-dependent pathway through the NPC that could involve local rearrangements to accommodate the hydrophilic domains of INM proteins. Our data suggest that these alterations in NPC structure may involve Nup170 and are consistent with a model in which Nup170 contributes to NPC gating³¹.

METHODS

Yeast strains. See Supplementary Table 2 for a list of all yeast strains used in this study and Supplementary Methods for their production. Yeast strains were grown at 30 °C, unless otherwise indicated, in YP (1% yeast extract, 2%

bactopeptone, 0.05% adenine) supplemented with either glucose (YPD) or raffinose (YPR) at a final concentration of 2%, or in complete synthetic medium lacking the appropriate amino acid.

Plasmids. See Supplementary Information for details regarding plasmid construction.

Nuclear envelope targeting assay. All micrographs of live cells were acquired using a Leica DM IRBE confocal microscope, or Zeiss LSM spinning disk confocal microscope. Yeast strains expressing YFP fusions under the control of a galactose inducible promoter (*GAL1*) were grown to an optical density of 0.5 at 600 nm in YPR. Galactose was added to a final concentration of 3% to induce expression of the YFP fusions. Images were acquired between 30 and 60 min after induction. In the case of temperature-sensitive mutants, cell cultures were first shifted to 34 °C before addition of galactose. To evaluate the recovery of the *mtr1-1* and *rna1-1* strains, glucose was added (final concentration, 4%) to repress new gene transcription, and images were acquired after 1 h at 25 °C. For strains CPL179–180, which express a cold-sensitive *ndc1-1* (ref. 20) allele, cells were first shifted to 15 °C before galactose induction.

Immunoelectron microscopy. Yeast strains expressing Heh1–YFP or Heh2–YFP were fixed and processed for immunoelectron microscopy as described in Supplementary Information.

In vitro binding assays. The production and purification of recombinant proteins and subsequent GST pull-down assays were performed as described³².

Received 8 June; accepted 11 July 2006.

Published online 23 August 2006.

- Gruenbaum, Y., Margalit, A., Goldman, R. D., Shumaker, D. K. & Wilson, K. L. The nuclear lamina comes of age. *Nature Rev. Mol. Cell Biol.* **6**, 21–31 (2005).
- Mounkes, L., Kozlov, S., Burke, B. & Stewart, C. L. The laminopathies: nuclear structure meets disease. *Curr. Opin. Genet. Dev.* **13**, 223–230 (2003).
- Suntharalingam, M. & Went, S. R. Peering through the pore: nuclear pore complex structure, assembly, and function. *Dev. Cell* **4**, 775–789 (2003).
- Soullam, B. & Worman, H. J. Signals and structural features involved in integral membrane protein targeting to the inner nuclear membrane. *J. Cell Biol.* **130**, 15–27 (1995).
- Ohba, T., Schirmer, E. C., Nishimoto, T. & Gerace, L. Energy- and temperature-dependent transport of integral proteins to the inner nuclear membrane via the nuclear pore. *J. Cell Biol.* **167**, 1051–1062 (2004).
- Mans, B. J., Anantharaman, V., Aravind, L. & Koonin, E. V. Comparative genomics, evolution and origins of the nuclear envelope and nuclear pore complex. *Cell Cycle* **3**, 1612–1637 (2004).
- Rodríguez-Navarro, S., Igual, J. C. & Pérez-Ortín, J. E. SRC1: an intron-containing yeast gene involved in sister chromatid segregation. *Yeast* **19**, 43–54 (2002).
- Macara, I. G. Transport into and out of the nucleus. *Microbiol. Mol. Biol. Rev.* **65**, 570–594 (2001).
- Fried, H. & Kutay, U. Nucleocytoplasmic transport: taking an inventory. *Cell. Mol. Life Sci.* **60**, 1659–1688 (2003).
- Weis, K. Regulating access to the genome: nucleocytoplasmic transport throughout the cell cycle. *Cell* **112**, 441–451 (2003).
- Kadowaki, T. et al. Isolation and characterization of *Saccharomyces cerevisiae* mRNA transport-defective (*mtr*) mutants. *J. Cell Biol.* **126**, 649–659 (1994).
- Kalderon, D., Richardson, W. D., Markham, A. F. & Smith, A. E. Sequence requirements for nuclear location of simian virus 40 large-T antigen. *Nature* **311**, 33–38 (1984).
- Yano, R., Oakes, M. L., Tabb, M. M. & Nomura, M. Yeast Srp1p has homology to armadillo/plakoglobin/β-catenin and participates in apparently multiple nuclear functions including the maintenance of the nucleolar structure. *Proc. Natl Acad. Sci. USA* **91**, 6880–6884 (1994).
- Irvine, M. K. & Went, S. R. A nuclear export signal in Kap95p is required for both recycling the import factor and interaction with the nucleoporin GLFG repeat regions of Nup116p and Nup100p. *J. Cell Biol.* **137**, 797–811 (1997).
- Leslie, D. M., Grill, B., Rout, M. P., Wozniak, R. W. & Aitchison, J. D. Kap121p-mediated nuclear import is required for mating and cellular differentiation in yeast. *Mol. Cell. Biol.* **22**, 2544–2555 (2002).
- Ryan, K. J. & Went, S. R. The nuclear pore complex: a protein machine bridging the nucleus and cytoplasm. *Curr. Opin. Cell Biol.* **12**, 361–371 (2000).
- Nehrbass, U., Rout, M. P., Maguire, S., Blobel, G. & Wozniak, R. W. The yeast nucleoporin Nup188p interacts genetically and physically with the core structures of the nuclear pore complex. *J. Cell Biol.* **133**, 1153–1162 (1996).
- Aitchison, J. D., Rout, M. P., Marelli, M., Blobel, G. & Wozniak, R. W. Two novel related yeast nucleoporins Nup170p and Nup157p: complementation with the vertebrate homologue Nup155p and functional interactions with the yeast nuclear pore-membrane protein Pom152p. *J. Cell Biol.* **131**, 1133–1148 (1995).
- Miao, M., Ryan, K. J. & Went, S. R. The integral membrane protein pom34p functionally links nucleoporin subcomplexes. *Genetics* **172**, 1441–1457 (2006).
- Chial, H. J., Rout, M. P., Giddings, T. H. & Winey, M. *Saccharomyces cerevisiae* Ndc1p is a shared component of nuclear pore complexes and spindle pole bodies. *J. Cell Biol.* **143**, 1789–1800 (1998).

21. Vaughan, A. *et al.* Both emerin and lamin C depend on lamin A for localization at the nuclear envelope. *J. Cell Sci.* **114**, 2577–2590 (2001).
 22. Wu, W., Lin, F. & Worman, H. J. Intracellular trafficking of MAN1, an integral protein of the nuclear envelope inner membrane. *J. Cell Sci.* **115**, 1361–1371 (2002).
 23. Ostlund, C., Sullivan, T., Stewart, C. L. & Worman, H. J. Dependence of diffusional mobility of integral inner nuclear membrane proteins on A-type lamins. *Biochemistry* **45**, 1374–1382 (2006).
 24. Horton, P. & Nakai, K. Better prediction of protein cellular localization sites with the k nearest neighbors classifier. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **5**, 147–152 (1997).
 25. Beilharz, T., Egan, B., Silver, P. A., Hofmann, K. & Lithgow, T. Bipartite signals mediate subcellular targeting of tail-anchored membrane proteins in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **278**, 8219–8223 (2003).
 26. Booth, J. W., Belanger, K. D., Sannella, M. I. & Davis, L. I. The yeast nucleoporin Nup2p is involved in nuclear export of importin α /Srp1p. *J. Biol. Chem.* **274**, 32360–32367 (1999).
 27. Gilchrist, D., Mykytko, B. & Rexach, M. Accelerating the rate of disassembly of karyopherin-cargo complexes. *J. Biol. Chem.* **277**, 18161–18172 (2002).
 28. Hood, J. K., Casolari, J. M. & Silver, P. A. Nup2p is located on the nuclear side of the nuclear pore complex and coordinates Srp1p/importin- α export. *J. Cell Sci.* **113**, 1471–1480 (2000).
 29. Solsbacher, J., Maurer, P., Vogel, F. & Schlenstedt, G. Nup2p, a yeast nucleoporin, functions in bidirectional transport of importin- α . *Mol. Cell. Biol.* **20**, 8468–8479 (2000).
 30. Sakseena, S., Summers, M. D., Burks, J. K., Johnson, A. E. & Braunagel, S. C. Importin- α 16 is a translocon-associated protein involved in sorting membrane proteins to the nuclear envelope. *Nature Struct. Mol. Biol.* **13**, 500–508 (2006).
 31. Shulga, N. & Goldfarb, D. S. Binding dynamics of structural nucleoporins govern nuclear pore complex permeability and may mediate channel gating. *Mol. Cell. Biol.* **23**, 534–542 (2003).
 32. Makhnevych, T., Lusk, C. P., Anderson, A. M., Aitchison, J. D. & Wozniak, R. W. Cell cycle regulated transport controlled by alterations in the nuclear pore complex. *Cell* **115**, 813–823 (2003).
- Supplementary Information** is linked to the online version of the paper at www.nature.com/nature. A summary figure is also included.
- Acknowledgements** We are grateful to S. Wente, R. Wozniak, X. Zhao, M. Winey and K. Belanger for yeast strains and plasmids, and M. Rout and J. Novatt for the anti-GFP antibody and Kap95. We also thank A. North for help with the spinning disk confocal microscope, R. Peters and L. Gerace for discussions, and E. Wren for helping us launch this project. We are especially indebted to H. Shio for electron micrograph technical support. This work was supported by an NIH fellowship (to M.C.K.) and the Howard Hughes Medical Institute (to G.B. and C.P.L.).
- Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.C.K. (mking@rockefeller.edu) or C.P.L. (plusk@rockefeller.edu).

LETTERS

The association of GRB 060218 with a supernova and the evolution of the shock wave

S. Campana¹, V. Mangano², A. J. Blustin³, P. Brown⁴, D. N. Burrows⁴, G. Chincarini^{1,5}, J. R. Cummings^{6,7}, G. Cusumano², M. Della Valle^{8,9}, D. Malesani¹⁰, P. Mészáros^{4,11}, J. A. Nousek⁴, M. Page³, T. Sakamoto^{6,7}, E. Waxman¹², B. Zhang¹³, Z. G. Dai^{13,14}, N. Gehrels⁶, S. Immler⁶, F. E. Marshall⁶, K. O. Mason¹⁵, A. Moretti¹, P. T. O'Brien¹⁶, J. P. Osborne¹⁶, K. L. Page¹⁶, P. Romano¹, P. W. A. Roming⁴, G. Tagliaferri¹, L. R. Cominsky¹⁷, P. Giommi¹⁸, O. Godet¹⁶, J. A. Kennea⁴, H. Krimm^{6,19}, L. Angelini⁶, S. D. Barthelmy⁶, P. T. Boyd⁶, D. M. Palmer²⁰, A. A. Wells¹⁶ & N. E. White⁶

Although the link between long γ -ray bursts (GRBs) and supernovae has been established^{1–4}, hitherto there have been no observations of the beginning of a supernova explosion and its intimate link to a GRB. In particular, we do not know how the jet that defines a γ -ray burst emerges from the star's surface, nor how a GRB progenitor explodes. Here we report observations of the relatively nearby GRB 060218 (ref. 5) and its connection to supernova SN 2006aj (ref. 6). In addition to the classical non-thermal emission, GRB 060218 shows a thermal component in its X-ray spectrum, which cools and shifts into the optical/ultraviolet band as time passes. We interpret these features as arising from the break-out of a shock wave driven by a mildly relativistic shell into the dense wind surrounding the progenitor⁷. We have caught a supernova in the act of exploding, directly observing the shock break-out, which indicates that the GRB progenitor was a Wolf-Rayet star.

GRB 060218 was detected with the Burst Alert Telescope (BAT) instrument⁸ onboard the Swift⁹ space mission on 18.149 February 2006 Universal Time⁵. The burst profile is unusually long with a T_{90} (the time interval containing 90% of the flux) of $2,100 \pm 100$ s (Fig. 1). The flux slowly rose to the peak at 431 ± 60 s (90% containment; times are measured from the BAT trigger time). Swift slewed autonomously to the newly discovered burst. The X-ray Telescope (XRT)¹⁰ found a bright source, which rose smoothly to a peak of ~ 100 counts s^{-1} (0.3–10 keV) at 985 ± 15 s. The X-ray flux then decayed exponentially with an e -folding time of $2,100 \pm 50$ s, followed around 10 ks later by a shallower power-law decay similar to that seen in typical GRB afterglows^{11,12} (Fig. 2). The UltraViolet/Optical Telescope (UVOT)¹³ found emission steadily brightening by a factor of 5–10 after the first detection, peaking in a broad plateau first in the ultraviolet (31.3 ± 1.8 ks at 188 nm) and later in the optical (39.6 $\pm$ 2.5 ks at 439 nm) parts of the spectrum. The light curves reached a minimum at about 200 ks, after which the ultraviolet light curves remained constant while a rebrightening was seen in the optical bands, peaking again at about 700–800 ks (Fig. 2).

Soon after the Swift discovery, low-resolution spectra of the optical afterglow and host galaxy revealed strong emission lines at a redshift of $z = 0.033$ (ref. 14). Spectroscopic indications of the presence of a rising supernova (designated SN 2006aj) were found three days after the burst^{6,15} with broad emission features consistent with a type Ic supernova (owing to a lack of hydrogen and helium lines).

The Swift instruments provided valuable spectral information. The high-energy spectra soften with time and can be fitted with (cut-off) power laws. This power-law component can be ascribed to the usual GRB jet and afterglow. The most striking feature, however, is the presence of a soft component in the X-ray spectrum that is present in the XRT data up to $\sim 10,000$ s. The blackbody component shows a marginally decreasing temperature ($kT \approx 0.17$ keV, where k is the Boltzmann constant), and a clear increase in luminosity with time, corresponding to an increase in the apparent emission radius from $R_{BB}^X = (5.2 \pm 0.5) \times 10^{11}$ cm to $R_{BB}^X = (1.2 \pm 0.1) \times 10^{12}$ cm (Fig. 3). During the rapid decay ($t \approx 7,000$ s), a blackbody component is still present in the data with a marginally cooler temperature ($kT = 0.10 \pm 0.05$ keV) and a comparable emission radius: $R_{BB}^X = (6.5^{+1.4}_{-4.4}) \times 10^{11}$ cm. In the optical/ultraviolet band at 9 hours (32 ks) the blackbody peak is still above the UVOT energy range. At 120 ks the peak of the blackbody emission is within the UVOT passband, and the inferred temperature and radius are $kT = 3.7^{+1.9}_{-0.9}$ eV and $R_{BB}^{UV} = 3.29^{+0.94}_{-0.93} \times 10^{14}$ cm, implying an expansion speed of $(2.7 \pm 0.8) \times 10^9$ cm s^{-1} . This estimate is consistent with what we would expect for a supernova and it is also consistent with the line broadening observed in the optical spectra.

The thermal components are the key to interpreting this anomalous GRB. The high temperature (two million degrees) of the thermal X-ray component suggests that the radiation is emitted by a shock-heated plasma. The characteristic radius of the emitting region, $R_{shell} \approx (E/at^4)^{1/3} \approx 5 \times 10^{12}$ cm (E is the GRB isotropic energy and a is the radiation density constant), corresponds to the radius of a blue supergiant progenitor. However, the lack of hydrogen lines

¹INAF—Osservatorio Astronomico di Brera, via E. Bianchi 46, I-23807 Merate (LC), Italy. ²INAF—Istituto di Astrofisica Spaziale e Fisica Cosmica di Palermo, via U. La Malfa 153, I-90146 Palermo, Italy. ³UCL Mullard Space Science Laboratory, Holmbury St. Mary, Dorking, Surrey RH5 6NT, UK. ⁴Department of Astronomy and Astrophysics, Pennsylvania State University, University Park, Pennsylvania 16802, USA. ⁵Università degli studi di Milano Bicocca, piazza delle Scienze 3, I-20126 Milano, Italy. ⁶NASA—Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. ⁷National Research Council, 2101 Constitution Avenue NW, Washington DC 20418, USA. ⁸INAF—Osservatorio Astrofisico di Arcetri, largo E. Fermi 5, I-50125 Firenze, Italy. ⁹Kavli Institute for Theoretical Physics, UC Santa Barbara, California 93106, USA. ¹⁰International School for Advanced Studies (SISSA-ISAS), via Beirut 2-4, I-34014 Trieste, Italy. ¹¹Department of Physics, Pennsylvania State University, University Park, Pennsylvania 16802, USA. ¹²Physics Faculty, Weizmann Institute, Rehovot 76100, Israel. ¹³Department of Physics, University of Nevada, Box 454002, Las Vegas, Nevada 89154-4002, USA. ¹⁴Department of Astronomy, Nanjing University, Nanjing, 210093, China. ¹⁵PPARC, Polaris House, North Star Avenue, Swindon SN2 1SZ, UK. ¹⁶Department of Physics and Astronomy, University of Leicester, University Road, Leicester LE1 7RH, UK. ¹⁷Department of Physics and Astronomy, Sonoma State University, Rohnert Park, California 94928-3609, USA. ¹⁸ASI Science Data Center, via G. Galilei, I-00044 Frascati (Roma), Italy. ¹⁹Universities Space Research Association, 10211 Wincopin Circle Suite 500, Columbia, Maryland 21044-3431, USA. ²⁰Los Alamos National Laboratory, PO Box 1663, Los Alamos, New Mexico 87545, USA.

in the supernova spectrum suggests a much more compact source. The large emission radius may be explained in this case by the existence of a massive stellar wind surrounding the progenitor, as is common for Wolf–Rayet stars. The thermal radiation is observed once the shock driven into the wind reaches a radius, $\sim R_{\text{shell}}$, where the wind becomes optically thin.

The characteristic variability time is $R_{\text{shell}}/c \approx 200$ s, consistent with the smoothness of the X-ray pulse and the rapid thermal X-ray flux decrease at the end of the pulse. We interpret this as providing, for the first time, a direct measurement of the shock break-out^{16,17} of the stellar envelope and the stellar wind (first investigated by Colgate¹⁸). The fact that R_{shell} is larger than R_{BB}^{X} suggests that the shock expands in a non-spherical manner, reaching different points on the R_{shell} sphere at different times. This may be due to a non-spherical explosion (such as the presence of a jet), or a non-spherical wind^{19,20}. In addition, the shock break-out interpretation provides us with a delay between the supernova explosion and the GRB start of $\lesssim 4$ ks (ref. 21; see Fig. 1).

As the shock propagates into the wind, it compresses the wind plasma into a thin shell. The mass of this shell may be inferred from the requirement that its optical depth be close to unity, $M_{\text{shell}} \approx 4\pi R_{\text{shell}}^2/\kappa \approx 5 \times 10^{-7} M_{\odot}$ ($\kappa \approx 0.34 \text{ cm}^2 \text{ g}^{-1}$ is the opacity). This implies that the wind mass-loss rate is $\dot{M} \approx M_{\text{shell}} v_{\text{wind}}/R_{\text{shell}} \approx 3 \times 10^{-4} M_{\odot} \text{ yr}^{-1}$, for a wind velocity $v_{\text{wind}} = 10^8 \text{ cm s}^{-1}$, typical for Wolf–Rayet stars. Because the thermal energy density behind a radiation-dominated shock is $aT^4 \approx 3\rho v_s^2$ (ρ is the wind density at R_{shell} and v_s the shock velocity) we have $\rho \approx 10^{-12} \text{ g cm}^{-3}$, which

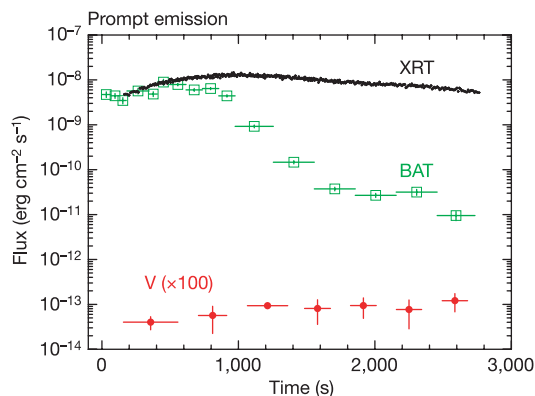


Figure 1 | Early Swift light curve of GRB 060218. GRB 060218 was discovered by the BAT when it came into the BAT field of view during a pre-planned slew. There is no emission at the GRB location up to $\sim 3,509$ s. Swift slewed again to the burst position and the XRT and UVOT began observing GRB 060218 159 s later. For each BAT point we converted the observed count rate to flux (15–150 keV band) using the observed spectra. The combined BAT and XRT spectra were fitted with a cut-off power law plus a blackbody, absorbed by interstellar matter in our Galaxy and in the host galaxy at redshift $z = 0.033$. The host galaxy column density is $N_{\text{H}}^z = 5.0 \times 10^{21} \text{ cm}^{-2}$ and that of our Galaxy is $(0.9\text{--}1.1) \times 10^{21} \text{ cm}^{-2}$. Errors are at 1σ significance. At a redshift $z = 0.033$ (corresponding to a distance of 145 Mpc with $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$) the isotropic equivalent energy, extrapolated to the 1–10,000-keV rest-frame energy band, is $E_{\text{iso}} = (6.2 \pm 0.3) \times 10^{49} \text{ erg}$. The peak energy in the GRB spectrum is at $E_p = 4.9_{-0.3}^{+0.4} \text{ keV}$. These values are consistent with the Amati correlation, suggesting that GRB060218 is not an off-axis event²⁶. This conclusion is also supported by the lack of achromatic rise behaviour of the light curve in the three Swift observation bands. The BAT fluence is dominated by soft X-ray photons and this burst can be classified as an X-ray flash²⁷. A V-band light curve is shown with red filled circles. For clarity the V flux has been multiplied by a factor of 100. Magnitudes have been converted to fluxes using standard UVOT zero points and multiplying the specific flux by the filter Full Width at Half Maximum (FWHM). Gaps in the light curve are due to the automated periodic change of filters during the first observation of the GRB.

implies that the shock must be (mildly) relativistic, $v_s \approx c$. This is similar to GRB 980425/SN 1998bw, where the ejection of a mildly relativistic shell with energy of $\approx 5 \times 10^{49} \text{ erg}$ is believed to have powered radio^{22–24} and X-ray emission⁷.

The optical–ultraviolet emission observed at an early time of $t \lesssim 10^4$ s may be accounted for as the low-energy tail of the thermal X-ray emission produced by the (radiation) shock driven into the wind. At a later time, the optical–ultraviolet emission is well above that expected from the (collisionless) shock driven into the wind. This emission is most probably due to the expanding envelope of the star, which was heated by the shock passage to a much higher temperature. Initially, this envelope is hidden by the wind. As the star and wind expand, the photosphere propagates inward, revealing shocked stellar plasma. As the star expands, the radiation temperature decreases and the apparent radius increases (Fig. 3). The radius inferred at the peak of the ultraviolet emission, $R_{\text{BB}}^{\text{UV}} \approx 3 \times 10^{14} \text{ cm}$, implies that emission is arising from the outer $\sim 4\pi(R_{\text{BB}}^{\text{UV}})^2/\kappa \approx 10^{-3} M_{\odot}$ shell at the edge of the shocked star. As the photosphere rapidly cools, this component of the emission fades. The ultraviolet light continues to plummet as cooler temperatures allow elements to recombine and line blanketing to set in, while radioactive decay causes the optical

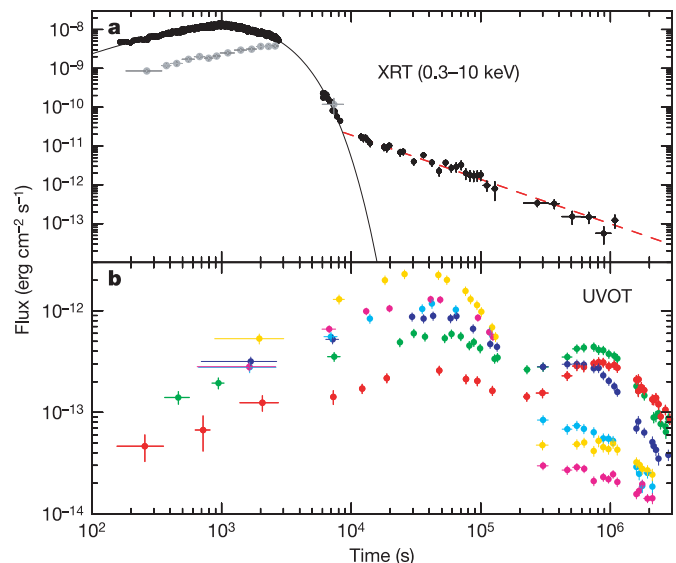


Figure 2 | Long-term Swift light curve of GRB 060218. **a**, The XRT light curve (0.3–10 keV) is shown with open black circles. Count-rate-to-flux conversion factors were derived from time-dependent spectral analysis. We also plotted (filled grey circles) the contribution to the 0.3–10-keV flux by the blackbody component. Its percentage contribution is increasing with time, becoming dominant at the end of the exponential (or steep power-law) decay. At about 10,000 s the light curve breaks to a shallower power-law decay (dashed red line) with an index of -1.2 ± 0.1 , characteristic of typical GRB afterglows. This classical afterglow can be naturally accounted for by a shock driven into the wind by a shell with kinetic energy $E_{\text{shell}} \approx 10^{49} \text{ erg}$. The t^{-1} flux decline is valid at the stage where the shell is being decelerated by the wind with the deceleration phase beginning at $t_{\text{dec}} \lesssim 10^4$ s for $\dot{M} \approx 10^{-4} (v_{\text{wind}}/10^8) M_{\odot} \text{ yr}^{-1}$ (where v_{wind} is in units of cm s^{-1}), consistent with the mass-loss rate inferred from the thermal X-ray component. Error bars are 1σ . **b**, The UVOT light curve. Filled circles of different colours represent different UVOT filters: red, V (centred at 544 nm); green, B (439 nm); dark blue, U (345 nm); light blue, UW1 (251 nm); magenta, UW2 (217 nm); and yellow, UW3 (188 nm). Specific fluxes have been multiplied by their FWHM widths (75, 98, 88, 70, 51 and 76 nm, respectively). Data have been rebinned to increase the signal-to-noise ratio. The ultraviolet band light curve peaks at about 30 ks owing to the shock break-out from the outer stellar surface and the surrounding dense stellar wind, while the optical band peaks at about 800 ks owing to radioactive heating in the supernova ejecta.

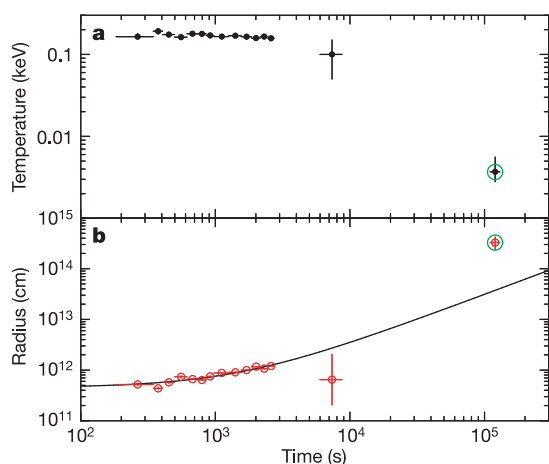


Figure 3 | Evolution of the soft thermal component temperature and radius. **a**, Evolution of the temperature of the soft thermal component. The joint BAT and XRT spectrum has been fitted with a blackbody component plus a (cut-off) power-law in the first $\sim 3,000$ s (see also the legend of Fig. 1). The last point (circled in green) comes from a fit to the six UVOT filters, assuming a blackbody model with Galactic reddening, $E(B - V) = 0.14$, and host galaxy reddening. This reddening has been determined by fitting the Rayleigh–Jeans tail of the blackbody emission at 32 ks (9 hours). The data require an intrinsic $E(B - V) = 0.20 \pm 0.03$ (assuming a Small Magellanic Cloud reddening law²⁸). Error bars are 1σ . **b**, Evolution of the radius of the soft thermal component. The last point (circled in green) comes from the fitting of UVOT data. The continuous line represents a linear fit to the data.

light to begin rising to the primary maximum normally seen in supernova light curves (Fig. 2).

Because the wind shell is clearly larger than the progenitor radius, we infer that the star radius is definitely smaller than 5×10^{12} cm. Assuming a linear expansion at the beginning (owing to light travel-time effects) we can estimate a star radius of $R_{\text{star}} \approx (4 \pm 1) \times 10^{11}$ cm. This is smaller than the radius of the progenitors of type II supernovae, like blue supergiants (4×10^{12} cm for SN 1987A, ref. 25) or red supergiants (3×10^{13} cm). Our results unambiguously indicate that the progenitor of GRB 060218/SN 2006aj was a compact massive star, most probably a Wolf–Rayet star.

Received 13 March; accepted 10 May 2005.

1. Woosley, S. E. Gamma-ray bursts from stellar mass accretion disks around black holes. *Astrophys. J.* **405**, 273–277 (1993).
2. Paczyński, B. Are gamma-ray bursts in star-forming regions? *Astrophys. J.* **494**, L45–L48 (1993).
3. MacFadyen, A. I. & Woosley, S. E. Collapsars: gamma-ray bursts and

- explosions in ‘failed supernovae’. *Astrophys. J.* **524**, 262–289 (1999).
4. Galama, T. J. et al. An unusual supernova in the error box of the γ -ray burst of 25 April 1998. *Nature* **395**, 670–672 (1998).
5. Cusumano, G. et al. GRB060218: Swift-BAT detection of a possible burst. *GCN Circ.* **4775** (2006).
6. Masetti, N. et al. GRB060218: VLT spectroscopy. *GCN Circ.* **4803** (2006).
7. Waxman, E. Does the detection of X-ray emission from SN1998bw support its association with GRB980425? *Astrophys. J.* **605**, L97–L100 (2004).
8. Barthelmy, S. D. et al. The Burst Alert Telescope (BAT) on the SWIFT Midex Mission. *Space Sci. Rev.* **120**, 143–164 (2005).
9. Gehrels, N. et al. The Swift gamma ray burst mission. *Astrophys. J.* **611**, 1005–1020 (2004).
10. Burrows, D. N. et al. The Swift X-Ray Telescope. *Space Sci. Rev.* **120**, 165–195 (2005).
11. Tagliaferri, G. et al. An unexpectedly rapid decline in the X-ray afterglow emission of long γ -ray bursts. *Nature* **436**, 985–988 (2005).
12. O’Brien, P. T. et al. The early X-ray emission from GRBs. *Astrophys. J.* (submitted); preprint at (<http://arXiv.org/astro-ph/0601125>) (2006).
13. Roming, P. W. A. et al. The Swift Ultra-Violet/Optical Telescope. *Space Sci. Rev.* **120**, 95–142 (2005).
14. Mirabal, N. & Halpern, J. P. GRB060218: MDM Redshift. *GCN Circ.* **4792** (2006).
15. Pian, E. et al. An optical supernova associated with the X-ray flash XRF 060218. *Nature* doi:10.1038/nature05082 (this issue).
16. Ensmann, L. & Burrows, A. Shock breakout in SN1987A. *Astrophys. J.* **393**, 742–755 (1992).
17. Tan, J. C., Matzner, C. D. & McKee, C. F. Trans-relativistic blast waves in supernovae as gamma-ray burst progenitors. *Astrophys. J.* **551**, 946–972 (2001).
18. Colgate, S. A. Prompt gamma rays and X-rays from supernovae. *Can. J. Phys.* **46**, 476 (1968).
19. Mazzali, P. A. et al. An asymmetric, energetic type Ic supernova viewed off-axis and a link to gamma-ray bursts. *Science* **308**, 1284–1287 (2005).
20. Leonard, D. C. et al. A non-spherical core in the explosion of supernova SN2004dj. *Nature* **440**, 505–507 (2006).
21. Norris, J. P. & Bonnell, J. T. How can the SN-GRB time delay be measured? *AIP Conf. Proc.* **727**, 412–415 (2004).
22. Kulkarni, S. R. et al. Radio emission from the unusual supernova 1998bw and its association with the gamma-ray burst of 25 April 1998. *Nature* **395**, 663–669 (1998).
23. Waxman, E. & Loeb, A. A subrelativistic shock model for the radio emission of SN1998bw. *Astrophys. J.* **515**, 721–725 (1999).
24. Li, Z.-Y. & Chevalier, R. A. Radio supernova SN1998bw and its relation to GRB980425. *Astrophys. J.* **526**, 716–726 (1999).
25. Arnett, W. D. et al. Supernova 1987A. *Annu. Rev. Astron. Astrophys.* **27**, 629–700 (1989).
26. Amati, L. et al. GRB060218: $E_{p,1} - E_{\text{iso}}$ correlation. *GCN Circ.* **4846** (2006).
27. Heise, J., et al. in *Proceedings of ‘Gamma-Ray Bursts in the Afterglow Era’* (eds Costa, E., Frontera, F. & Hjorth, J.) 16–21 (Springer, Berlin/Heidelberg, 2001).
28. Pei, Y. C. Interstellar dust from the Milky Way to the Magellanic Clouds. *Astrophys. J.* **395**, 130–139 (1992).

Acknowledgements We acknowledge support from ASI, NASA and PPARC.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.C. (sergio.campana@brera.inaf.it).

An optical supernova associated with the X-ray flash XRF 060218

E. Pian^{1,2}, P. A. Mazzali^{1,2,3,4,5}, N. Masetti⁶, P. Ferrero⁷, S. Klose⁷, E. Palazzi⁶, E. Ramirez-Ruiz^{8,9}, S. E. Woosley⁹, C. Kouveliotou¹⁰, J. Deng^{2,4,5,11}, A. V. Filippenko¹², R. J. Foley¹², J. P. U. Fynbo¹³, D. A. Kann⁷, W. Li¹², J. Hjorth¹³, K. Nomoto^{2,4,5}, F. Patat¹⁴, D. N. Sauer^{1,2}, J. Sollerman^{13,15}, P. M. Vreeswijk^{16,17}, E. W. Guenther⁷, A. Levani^{2,18}, P. O'Brien¹⁹, N. R. Tanvir¹⁹, R. A. M. J. Wijers²⁰, C. Dumas¹⁷, O. Hainaut¹⁷, D. S. Wong¹², D. Baade¹⁴, L. Wang^{21,22}, L. Amati⁶, E. Cappellaro²³, A. J. Castro-Tirado²⁴, S. Ellison²⁵, F. Frontera^{6,26}, A. S. Fruchter²⁷, J. Greiner²⁸, K. Kawabata²⁹, C. Ledoux¹⁷, K. Maeda^{2,30}, P. Møller¹⁴, L. Nicastro⁶, E. Rol¹⁹ & R. Starling²⁰

Long-duration γ -ray bursts (GRBs) are associated with type Ic supernovae¹ that are more luminous than average^{2–5} and that eject material at very high velocities. Less-luminous supernovae were not hitherto known to be associated with GRBs, and therefore GRB-supernovae were thought to be rare events⁶. Whether X-ray flashes—analogs of GRBs, but with lower luminosities and fewer γ -rays—can also be associated with supernovae, and whether they are intrinsically ‘weak’ events or typical GRBs viewed off the axis of the burst⁷, is unclear. Here we report the optical discovery and follow-up observations of the type Ic supernova SN 2006aj associated with X-ray flash XRF 060218. Supernova 2006aj is intrinsically less luminous than the GRB-supernovae, but more luminous than many supernovae not accompanied by a GRB. The ejecta velocities derived from our spectra are intermediate between these two groups, which is consistent with the weakness of both the GRB output⁸ and the supernova radio flux⁹. Our data, combined with radio and X-ray observations^{8–10}, suggest that XRF 060218 is an intrinsically weak and soft event, rather than a classical GRB observed off-axis. This extends the GRB-supernova connection to X-ray flashes and fainter supernovae, implying a common origin. Events such as XRF 060218 are probably more numerous than GRB-supernovae.

The Burst Alert Telescope (BAT) onboard the Swift spacecraft detected XRF 060218 on 18 February 2006 at 03:34:30 UT (ref. 8). Its spectrum peaked near 5 keV, placing the burst in the XRF subgroup of GRBs. The optical counterpart of the burst was detected ~ 200 s later by the Swift Ultraviolet/Optical Telescope, and was subsequently observed by ground-based telescopes¹¹. The closeness of

the event¹² made XRF 060218 an ideal candidate for spectroscopic observations of a possible associated supernova.

We observed XRF 060218 with the European Southern Observatory’s (ESO) 8.2-m Very Large Telescope (VLT) and the University of California’s Lick Observatory Shane 3-m telescope (Lick) starting 21 February 2006. Supplementary Table 1 shows the log of the observations. Spectroscopy was performed nearly daily for seventeen days (see Supplementary Fig. 1). Broad absorption lines detected in our first spectrum resembled those of broad-lined type Ic supernovae, thus providing the first definite case of a supernova associated with an XRF¹³. To our knowledge, this is the earliest spectroscopy of a GRB-supernova, and one of the earliest for any supernova. From its early decline, we estimate that the contribution of the fading afterglow of XRF 060218 to the supernova emission is not significant at the epoch of our first spectrum^{11,12}.

The high-dispersion spectrum taken with the VLT Ultraviolet and Visual Echelle Spectrograph (UVES) near the epoch of supernova maximum exhibits several narrow emission and absorption lines. From the former we obtained an accurate measurement of the host-galaxy redshift, $z = 0.03342 \pm 0.00002$ (heliocentrically corrected), corresponding to a distance of ~ 140 Mpc (using a Hubble constant of $H_0 = 73 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_A = 0.72$, and $\Omega_m = 0.28$). We constrained the total extinction toward the supernova from the equivalent widths of the interstellar Na I D absorption lines¹⁴ to be $E(B - V) = 0.13 \pm 0.02$ mag (P.A.M., manuscript in preparation). The extinction is mainly due to our Galaxy, and its value is consistent with that derived using infrared dust maps¹⁵. We used this value to correct the light curve of SN 2006aj (Fig. 1).

It is interesting to compare the properties of SN 2006aj with those

¹Istituto Nazionale di Astrofisica, Trieste Astronomical Observatory, via G. B. Tiepolo 11, I-34131 Trieste, Italy. ²Kavli Institute for Theoretical Physics, University of California, Santa Barbara, California 93106-4030, USA. ³Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Strasse 1, D-85748 Garching, Germany. ⁴Department of Astronomy, School of Science, University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan. ⁵Research Center for the Early Universe, School of Science, University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan. ⁶Istituto Nazionale di Astrofisica, IASF, Bologna, Via P. Gobetti 101, I-40129 Bologna, Italy. ⁷Thüringer Landessternwarte Tautenburg, Sternwarte 5, D-07778 Tautenburg, Germany. ⁸Institute for Advanced Study, Einstein Drive, Princeton, New Jersey 08540, USA. ⁹Department of Astronomy and Astrophysics, University of California, Santa Cruz, California 95064, USA. ¹⁰NASA/MSFC, NSSTC, VP62, 320 Sparkman Drive, Huntsville, Alabama 35805, USA. ¹¹National Astronomical Observatories, Chinese Academy of Sciences, 20A Datun Road, Chaoyang District, Beijing 100012, China. ¹²Department of Astronomy, University of California, Berkeley, California 94720-3411, USA. ¹³Dark Cosmology Centre, Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, DK-2100 Copenhagen Ø, Denmark. ¹⁴European Southern Observatory, Karl-Schwarzschild-Strasse 2, D-85748 Garching bei München, Germany. ¹⁵Stockholm Observatory, Department of Astronomy, AlbaNova, 106 91 Stockholm, Sweden. ¹⁶Departamento de Astronomía, Universidad de Chile, Casilla 36-D, Santiago, Chile. ¹⁷European Southern Observatory, Alonso de Córdova 3107, Casilla 19001, Santiago 19, Chile. ¹⁸Centre for Astrophysics Research, University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK. ¹⁹X-Ray and Observational Astronomy Group, Department of Physics and Astronomy, University of Leicester, Leicester LE1 7RH, UK. ²⁰Astronomical Institute ‘Anton Pannekoek’, University of Amsterdam, Kruislaan 403, 1098 SJ Amsterdam, The Netherlands. ²¹Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA. ²²Purple Mountain Observatory, Chinese Academy of Sciences, 2 Beijing Xi Lu, Nanjing, Jiangsu 210008, China. ²³Istituto Nazionale di Astrofisica, Padova Astronomical Observatory, Vicolo dell’Osservatorio 5, I-35122 Padova, Italy. ²⁴Instituto de Astrofísica de Andalucía (IAA-CSIC), Apartado de Correos 3004, 18080 Granada, Spain. ²⁵Department of Physics and Astronomy, University of Victoria, 3800 Finnerty Road, Victoria, British Columbia, V8P 1A1, Canada. ²⁶Department of Physics, University of Ferrara, Polo Scientifico e Tecnologico, Edificio C, via Saragat 1, I-44100 Ferrara, Italy. ²⁷Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218, USA. ²⁸Max-Planck-Institut für extraterrestrische Physik, Giessenbachstrasse, D-85741 Garching, Germany. ²⁹Hiroshima Astrophysical Science Center, Hiroshima University, Hiroshima 739-8526, Japan. ³⁰Department of Earth Science and Astronomy, College of Arts and Sciences, University of Tokyo, Komaba 3-8-1, Meguro-ku, Tokyo 153-8902, Japan.

of other type Ic supernovae. The three well-observed, low-redshift GRB-supernovae (SN 1998bw, SN 2003dh and SN 2003lw) are strikingly similar. They are about 5–6 times more luminous and about 30 times more energetic than typical type Ic supernovae¹⁶. The peak luminosities and the kinetic energies of the GRB-supernovae differ by no more than 30%. At maximum light, SN 2006aj is dimmer than these supernovae by about a factor of two, but it is still a factor of 2–3 more luminous than other broad-lined type Ic supernovae not associated with GRBs and normal (narrow-lined) type Ic supernovae (Fig. 1).

Normal type Ic supernovae rise to a peak in approximately 10–12 days and have photospheric expansion velocities of $\sim 10,000 \text{ km s}^{-1}$ after about 10 days. Previously known GRB-supernovae showed a longer rise time (14–15 days) and had, at an epoch of about ten days, velocities of $\sim 25,000 \text{ km s}^{-1}$ (see Figs 1 and 2). If XRF 060218 and SN 2006aj occurred simultaneously, SN 2006aj rose as fast as normal type Ic supernovae, and also declined comparatively fast. At the same time, the photospheric expansion velocity derived from spectral modelling is intermediate between the GRB-supernovae and other type Ic supernovae, broad-lined or narrow-lined, that were not associated with GRBs (Fig. 2). Asymmetry in the supernova explosion may modify the observed luminosity with respect to the intrinsic one, depending on the orientation of the symmetry axis, by no more than 25% (ref. 17).

We conclude that SN 2006aj is intrinsically dimmer than the other three GRB-supernovae. In addition, it is associated with the softest (but not the weakest) of the four local events connected with supernovae⁸, and it has mildly relativistic ejecta^{8,9}, thus appearing to be an intermediate object between GRB-supernovae and other type Ic supernovae, both broad-lined and narrow-lined, not accompanied by a GRB.

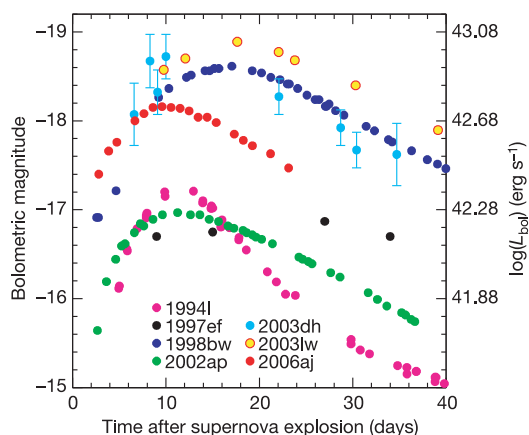


Figure 1 | Bolometric light curves of type Ic supernovae. We report, as a function of time, the luminosity and corresponding absolute magnitude of (1) the four spectroscopically identified supernovae associated with GRBs and XRFs, namely SN 1998bw (GRB 980425, $z = 0.0085$), SN 2003dh (GRB 030329, $z = 0.168$), SN 2003lw (GRB 031203, $z = 0.1055$), and SN 2006aj (XRF 060218, $z = 0.03342$); (2) of two broad-lined supernovae (not accompanied by a GRB), SN 1997ef and SN 2002ap; and (3) of the normal, intensively monitored SN 1994I. All represented supernovae are type Ic. The light curves, reported in their rest frame, have been constructed in the 3,000–24,000 Å range, taking into account the Galactic and, where appropriate, the host galaxy extinction^{16,25–28}. For SN 2006aj, we used the optical light curves obtained during our monitoring and the near-infrared data reported by ref. 29, and a total extinction value of $E(B - V) = 0.13 \text{ mag}$ (see text). We adopted the extinction curve of ref. 30 with $R_V = 3.1$. The galaxy contribution has also been subtracted where significant. The initial time has been assumed to coincide with the XRF detection time, 18 February 2006 at 03:34:30 UT. The systematic errors (about 0.2 mag) have been omitted, for clarity. Error bars are 1σ . The shape of the light curve of SN 2006aj is similar to that of SN 2002ap, as are the spectra¹⁹.

All together, these facts point to a substantial diversity between supernovae associated with GRBs and supernovae associated with XRFs. This diversity may be related to the masses of the exploding stars. In a companion paper, the parameters of the explosion are derived from models of the supernova optical light curves and spectra, and a relatively low initial mass, $20M_{\odot}$ (where $M_{\odot}$ is the mass of the sun), is proposed, evolving to a $3.3M_{\odot}$ CO star¹⁸. This mass is smaller than those estimated for the typical GRB-supernovae¹⁹.

GRBs and GRB-supernovae are aspherical sources. If XRF 060218 was a normal GRB viewed off-axis, the observed soft flux was emitted at large angles with respect to its jet axis. If the associated SN 2006aj is aspherical, then it is also probably seen off-axis. Alternatively, XRF 060218 may have been intrinsically soft, whether it was an aspherical explosion viewed on-axis or a spherical event. Various independent arguments, such as the chromatic behaviour of the multiwavelength counterpart of XRF 060218 (ref. 8), the absence of a late radio rebrightening⁹ and the compliance of XRF 060218 with the empirical correlation between peak energy and isotropic energy¹⁰, favour the latter possibility.

Together with the observation of other underluminous, relatively nearby XRFs and GRBs—GRB 980425 (ref. 2), XRF 030723 (refs 20, 21), XRF 020903 (ref. 22), and GRB 031203 (refs 23, 24), some definitely and some probably associated with supernovae—the properties of XRF 060218 suggest the existence of a population of events less luminous than ‘classical’ GRBs, but possibly much more numerous and with lower radio luminosities⁹. Indeed, these events may be the most abundant form of X- or γ -ray explosive transient in the Universe, but instrumental limits allow us to detect them only locally, so that several intrinsically subluminal bursts may remain undetected. The fraction of supernovae that are associated with GRBs or XRFs may be higher than currently thought.

By including this underluminous population and assuming no correction for possible collimation, which may vary from object to object, we obtain a local GRB rate of $110^{+180}_{-20} \text{ Gpc}^{-3} \text{ yr}^{-1}$, compared to $1 \text{ Gpc}^{-3} \text{ yr}^{-1}$ estimated from the cosmological events only (see Supplementary Information for details). In particular, for the detection threshold of Swift, we expect a few bursts per year within $z = 0.1$ and with luminosities as low as that of GRB 980425. The low-energy GRB population could be part of a continuum of explosion phenomena that mark the collapse of a stellar core, with normal supernovae at one end and classical GRBs at the other.

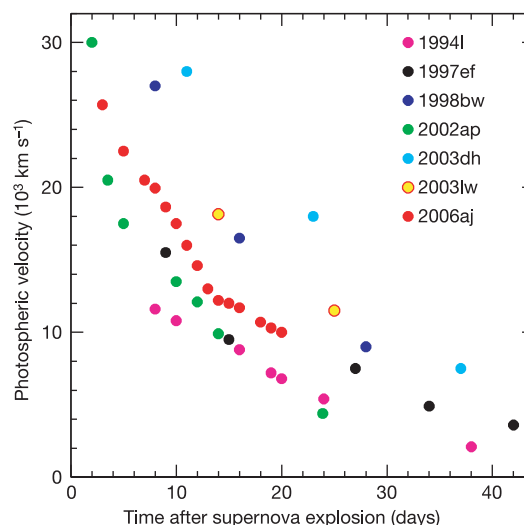


Figure 2 | Photospheric expansion velocities of type Ic supernovae. The time profiles of the expansion velocities of the same seven supernovae represented in Fig. 1 are reported. The velocities have been determined through models of the spectra at the various epochs^{16,18,25,26}.

Received 20 March; accepted 10 July 2006.

1. Filippenko, A. V. Optical spectra of supernovae. *Annu. Rev. Astron. Astrophys.* **35**, 309–355 (1997).
2. Galama, T. J. *et al.* An unusual supernova in the error box of the γ -ray burst of 25 April 1998. *Nature* **395**, 670–672 (1998).
3. Hjorth, J. *et al.* A very energetic supernova associated with the γ -ray burst of 29 March 2003. *Nature* **423**, 847–850 (2003).
4. Stanek, K. Z. *et al.* Spectroscopic discovery of the supernova 2003dh associated with GRB 030329. *Astrophys. J.* **591**, L17–L20 (2003).
5. Malesani, D. *et al.* SN 2003lw and GRB 031203: a bright supernova for a faint Gamma-Ray Burst. *Astrophys. J.* **609**, L5–L8 (2004).
6. Woosley, S. E. & Bloom, J. S. The supernova–gamma-ray-burst connection. *Annu. Rev. Astron. Astrophys.* **44**, 507–556 (2006).
7. Granot, J., Ramirez-Ruiz, E. & Perna, R. Afterglow observations shed new light on the nature of X-ray flashes. *Astrophys. J.* **630**, 1003–1014 (2005).
8. Campana, S. *et al.* The association of GRB 060218 with a supernova and the evolution of the shock wave. *Nature* doi:10.1038/nature04892 (this issue).
9. Soderberg, A. M. *et al.* Relativistic ejecta from X-ray flash XRF 060218 and the rate of cosmic explosions. *Nature* doi:10.1038/nature05087 (this issue).
10. Amati, L., Frontera, F., Guidorzi, C. & Montanari, E. GRB 060218: Ep,i–Eiso correlation. *GCN Circ.* **4846** (2006).
11. Modjaz, M. *et al.* Early-time photometry and spectroscopy of the fast evolving SN 2006aj associated with GRB 060218. *Astrophys. J.* **645**, L21–L24 (2006).
12. Mirabal, N., Halpern, J. P., An, D., Thorstensen, J. R. & Terndrup, D. M. GRB 060218A/SN 2006aj: a gamma-ray burst and prompt supernova at $z = 0.0335$. *Astrophys. J.* **643**, L99–L102 (2006).
13. Masetti, N., Palazzi, E., Pian, E. & Patat, F. GRB 060218: VLT spectroscopy. *GCN Circ.* **4803** (2006).
14. Guenther, E. W., Klose, S., Vreeswijk, P. M., Pian, E. & Greiner, J. GRB 060218/SN 2006aj, high resolution spectra. *GCN Circ.* **4863** (2006).
15. Schlegel, D. J., Finkbeiner, D. P. & Davis, M. Maps of dust infrared emission for use in estimation of reddening and cosmic microwave background radiation foregrounds. *Astrophys. J.* **500**, 525–553 (1998).
16. Mazzali, P. A. *et al.* Models for the type Ic hypernova SN 2003lw associated with GRB 031203. *Astrophys. J.* **645**, 1323–1330 (2006).
17. Maeda, K., Mazzali, P. A. & Nomoto, K. Optical emission from aspherical supernovae and the hypernova SN 1998bw. *Astrophys. J.* **645**, 1331–1344 (2006).
18. Mazzali, P. A. *et al.* A neutron-star-driven X-ray flash associated with supernova SN 2006aj. *Nature* doi:10.1038/nature05081 (this issue).
19. Woosley, S. E. & Heger, A. *et al.* The progenitor stars of gamma-ray bursts. *Astrophys. J.* **637**, 914–921 (2006).
20. Fynbo, J. P. U. *et al.* On the afterglow of the X-ray flash of 2003 July 23: photometric evidence for an off-axis gamma-ray burst with an associated supernova? *Astrophys. J.* **609**, 962–971 (2004).
21. Tominaga, N. *et al.* Supernova light-curve models for the bump in the optical counterpart of X-ray flash 030723. *Astrophys. J.* **612**, L105–L108 (2004).
22. Soderberg, A. M. *et al.* An HST search for supernovae accompanying X-ray flashes. *Astrophys. J.* **627**, 877–887 (2005).
23. Sazonov, S. Y., Lutovinov, A. A. & Sunyaev, R. A. An apparently normal gamma-ray burst with an unusually low luminosity. *Nature* **430**, 646–648 (2004).
24. Soderberg, A. M. *et al.* The sub-energetic gamma-ray burst GRB 031203 as a cosmic analogue to the nearby GRB 980425. *Nature* **430**, 648–650 (2004).
25. Sauer, D. N. *et al.* The properties of the ‘standard’ type Ic supernova 1994I from spectral models. *Mon. Not. R. Astron. Soc.* **369**, 1939–1948 (2006).
26. Mazzali, P. A. *et al.* Properties of two hypernovae entering the nebular phase: SN 1997ef and SN 1997dq. *Astrophys. J.* **614**, 858–863 (2004).
27. Tomita, H. *et al.* The optical/near-infrared light curves of SN 2002ap for the first 1.5 years after discovery. *Astrophys. J.* **644**, 400–408 (2006).
28. Deng, J., Tominaga, N., Mazzali, P. A., Maeda, K. & Nomoto, K. On the light curve and spectrum of SN 2003dh separated from the optical afterglow of GRB 030329. *Astrophys. J.* **624**, 898–905 (2005).
29. Cobb, B. E., Bailyn, C. D., van Dokkum, P. G. & Natarajan, P. SN 2006aj and the nature of low-luminosity gamma-ray bursts. *Astrophys. J.* **645**, L113–L116 (2006).
30. Cardelli, J. A., Clayton, G. C. & Mathis, J. S. The relationship between infrared, optical, and ultraviolet extinction. *Astrophys. J.* **345**, 245–256 (1989).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work is based on data collected by the GRACE consortium with ESO Paranal telescopes. The ESO staff astronomers at Paranal are acknowledged for their professional assistance. We are grateful to S. R. Kulkarni, M. Modjaz, A. Rau, and S. Savaglio for helpful interactions and to R. Wilman for allowing us to implement our Target-of-Opportunity programme with the VLT during his scheduled observing time. We thank S. Barthelmy for providing information about the Swift/BAT performance. This work has benefited from collaboration within the EU FP5 Research Training Network ‘Gamma-Ray Bursts: an Enigma and a Tool’. IRAF is distributed by the National Optical Astronomy Observatories, which are operated by the Association of Universities for Research in Astronomy, Inc, under contract to the National Science Foundation (NSF). A.V.F.’s group at the University of California, Berkeley, is supported by the NSF and by the TABASGO Foundation.

Author Contributions E. Pian, N.M., P.F., S.K., E. Palazzi, A.V.F., R.J.F., W.L., F.P., P.M.V., E.W.G., C.D., O.H., D.S.W., D.B., L.W., S.E. and C.L. organized the observations and were responsible for data acquisition, reduction and analysis. P.A.M., E.R.-R., S.E.W., J.D., K.N., D.N.S. and K.M. contributed to the interpretation and discussion of the data. J.P.U.F., D.A.K., J.H., J.S., A.L., P.O.B., L.A., E.C., A.J.C.-T., F.F., A.S.F., J.G., K.K., P.M., L.N. and E.R. provided expertise on specific aspects of the data presentation and discussion. E. Pian, P.A.M., E.R.-R., S.E.W., C.K., K.N., N.R.T., R.A.M.J.W., E.C. and R.S. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to E.P. (pian@oats.inaf.it).

LETTERS

Relativistic ejecta from X-ray flash XRF 060218 and the rate of cosmic explosions

A. M. Soderberg¹, S. R. Kulkarni¹, E. Nakar², E. Berger³, P. B. Cameron¹, D. B. Fox⁴, D. Frail⁵, A. Gal-Yam¹, R. Sari¹, S. B. Cenko⁶, M. Kasliwal¹, R. A. Chevalier⁷, T. Piran⁸, P. A. Price⁹, B. P. Schmidt¹⁰, G. Pooley¹¹, D.-S. Moon⁶, B. E. Penprase¹², E. Ofek¹, A. Rau¹, N. Gehrels¹³, J. A. Nousek⁴, D. N. Burrows⁴, S. E. Persson³ & P. J. McCarthy³

Over the past decade, long-duration γ -ray bursts (GRBs)—including the subclass of X-ray flashes (XRFs)—have been revealed^{1–3} to be a rare variety of type Ibc supernova. Although all these events result from the death of massive stars, the electromagnetic luminosities of GRBs and XRFs exceed those of ordinary type Ibc supernovae by many orders of magnitude. The essential physical process that causes a dying star to produce a GRB or XRF, and not just a supernova, is still unknown. Here we report radio and X-ray observations of XRF 060218 (associated⁴ with supernova SN 2006aj), the second-nearest^{5,6} GRB identified until now. We show that this event is a hundred times less energetic but ten times more common than cosmological GRBs. Moreover, it is distinguished from ordinary type Ibc supernovae by the presence of 10^{48} erg coupled to mildly relativistic ejecta, along with a central engine (an accretion-fed, rapidly rotating compact source) that produces X-rays for weeks after the explosion. This suggests that the production of relativistic ejecta is the key physical distinction between GRBs or XRFs and ordinary supernovae, while the nature of the central engine (black hole or magnetar) may distinguish typical bursts from low-luminosity, spherical events like XRF 060218.

On 2006 February 18.15 UT, the Burst Alert Telescope (BAT), a hard X-ray detector aboard the Swift satellite, detected⁵ an exceedingly long-duration ($\Delta t \approx 2,000$ s) transient. Within 153 s of the γ -ray trigger, the on-board X-ray Telescope (XRT) and Ultra-Violet Optical Telescope (UVOT) identified⁵ a counterpart coincident⁶ with a dwarf galaxy at $z = 0.0335$. The XRT and BAT data show⁵ that the event peaked at a photon energy of 4.9 keV, thus classifying this transient as an X-ray flash: XRF 060218. Distinguished⁷ by their soft X-ray dominated spectrum (peak energy, $E_p \approx 25$ keV versus 250 keV), the subclass of XRFs are otherwise similar (see ref. 8 and references therein) to GRBs in their observational properties.

Using the Very Large Array (VLA), we discovered a radio source at $\alpha(\text{J2000}) = 03^{\text{h}} 21^{\text{m}} 39.68^{\text{s}}$ and $\delta(\text{J2000}) = 16^{\circ} 52' 01.82''$ (± 0.02 arcsec in each axis), coincident with the UVOT position. Our monitoring of the radio source showed a power-law decay with $\alpha \approx -0.8$ through $t \approx 22$ days (Table 1), similar to the decay of afterglows seen from GRBs; here $F_\nu \propto t^\alpha$ is the spectral flux density. Over the same period the XRT undertook intensive observations of the source in the X-ray band (0.3–10 keV). We find the X-ray spectral flux density, $F_{\nu, X} \propto \nu_X^\beta$, is fitted by $\beta_X = -2.2 \pm 0.2$ with an absorbing hydrogen column density of $N_H = 3.9 \pm 0.4 \times 10^{21} \text{ cm}^{-2}$, consistent with previously reported^{5,9} values.

Separately, we observed the source with the Advanced CCD Imaging Spectrometer (ACIS) instrument aboard the Chandra X-ray Observatory (CXO). These observations began on 2006 February 26.78 and March 7.55 UT ($t \approx 8.8$ and 17.4 days) and lasted about 20 and 30 ks, respectively. The measured count rates are $(1.9 \pm 0.3) \times 10^{-3}$ and $(1.3 \pm 0.3) \times 10^{-3} \text{ s}^{-1}$, respectively. Using the XRT model parameters stated above we derive $F_X = (4.5 \pm 1.4) \times 10^{-14}$ and $(2.8 \pm 0.9) \times 10^{-14} \text{ erg cm}^{-2} \text{ s}^{-1}$ for the unabsorbed flux values. The XRT-CXO data spanning the range from a few minutes to 17 days are well fitted by a simple power-law decay model with temporal index, $\alpha_X = -1.1$.

XRF 060218 is most interesting because it is nearby, at a distance $d \approx 145$ Mpc. Indeed it is second only to GRB 980425/SN 1998bw¹, at just 36 Mpc. Similar to GRB 980425, XRF 060218 is also associated⁴ with a type Ic supernova explosion, SN 2006aj. The isotropic prompt energy release⁵ $E_{\gamma, \text{iso}} = (6.2 \pm 0.3) \times 10^{49} \text{ erg}$, is at least a hundred times fainter than typical GRBs but comparable to another nearby event, GRB 031203^{10,11} ($z = 0.106$). Similarly, the radio and X-ray luminosities are 10^3 and 10^2 times fainter than those of cosmological GRBs, respectively.

Radio observations directly probe the ejecta and environments of stellar explosions because the blastwave (velocity v) shocks the circumstellar medium and accelerates relativistic electrons that give rise to radio synchrotron emission. For radio sources dominated by synchrotron self-absorption, the brightness temperature is $T_B \lesssim 4 \times 10^{10} \text{ K}$. As can be seen from Fig. 1, at day 5 the radio emission peaks between 1.4 GHz and 4.9 GHz. Applying the basic equipartition analysis (see ref. 2) we find, at this epoch, that the radius of the radio-emitting region is $r \approx 3 \times 10^{16} \text{ cm}$, the ejecta kinetic energy is $E_K \approx 2 \times 10^{48} \text{ erg}$ and the circumburst particle density is $n \approx 5 \text{ cm}^{-3}$. The blastwave thus expands with a Lorentz factor of $\Gamma = (1 - \beta^2)^{-1/2} \approx 2.3$; here $\beta \equiv v/c$.

The early, steady decay of the radio emission indicates¹² that it cannot be attributed to a collimated jet directed away from our line-of-sight. Moreover, on a timescale of $t_{\text{NR}} \approx 7.3(E_{K,48}/n_0)^{1/3}$ days, the blastwave becomes¹³ sub-relativistic ($\Gamma\beta < 1$), at which point it effectively assumes spherical geometry, even if the initial explosion was biconical. Independently, noting the absence of a 'jet break' in the radio light-curve (to 22 days) and applying the standard formulation¹⁴ we find the opening angle, $\theta_j \gtrsim 1.4$ radians. Thus, on several grounds, the radio data argue for a quasi-spherical ejecta with 10^{48} erg coupled to mildly relativistic material. In addition, our

¹Caltech Optical Observatories 105-24, ²Theoretical Astrophysics 130-33, ³Space Radiation Laboratory 220-47, California Institute of Technology, Pasadena, California 91125, USA. ⁴Carnegie Observatories, 813 Santa Barbara Street, Pasadena, California 91101, USA. ⁵Department of Astronomy, Pennsylvania State University, University Park, Pennsylvania 16802, USA. ⁶National Radio Astronomy Observatory, PO Box 0, Socorro, New Mexico 87801, USA. ⁷Department of Astronomy, University of Virginia, PO Box 3818, Charlottesville, Virginia 22903, USA. ⁸Racah Institute of Physics, Hebrew University, Jerusalem 91904, Israel. ⁹Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, Hawaii 96822, USA. ¹⁰RSAA, ANU, Mt Stromlo Observatory, via Cotter Road, Weston Creek, Australian Capital Territory 2611, Australia. ¹¹Mullard Radio Astronomy Observatory, Cavendish Laboratory, Cambridge CB3 0HE, UK. ¹²Pomona College Dept. of Physics & Astronomy, 610 North College Ave, Claremont, California 91711, USA. ¹³NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA.

Table 1 | Radio observations made with the VLA and the Ryle telescope†

Epoch (UT)	Δt (days)	$F_{1.43}$ (μJy)	$F_{4.86}$ (μJy)	$F_{8.46}$ (μJy)	$F_{15.0}$ (μJy)	$F_{22.5}$ (μJy)
2006 Feb 20.02	1.87	-	78 ± 70	453 ± 77	-	-
2006 Feb 21.14	3.00	-	-	381 ± 60	-	250 ± 52
2006 Feb 21.77†	3.62	-	-	-	350 ± 350	-
2006 Feb 21.97	3.83	-	287 ± 56	269 ± 40	-	-
2006 Feb 22.99	4.85	25 ± 25	328 ± 61	280 ± 47	-	-
2006 Feb 25.12	6.97	134 ± 145	80 ± 47	164 ± 39	46 ± 141	-
2006 Feb 26.09	7.94	-	32 ± 32	30 ± 30	-	-
2006 Feb 28.10	9.95	-	-	39 ± 25	-	-
2006 Mar 2.23	12.08	70 ± 70	-	-	-	-
2006 Mar 3.03	12.88	-	-	15 ± 15	-	-
2006 Mar 6.89	16.74	-	-	75 ± 13	-	-
2006 Mar 10.01	19.86	-	-	48 ± 14	-	-
2006 Mar 12.11	21.96	-	-	87 ± 39	-	-
2006 Mar 15.04	24.91	-	-	20 ± 20	-	-
2006 Mar 20.86	30.71	-	-	32 ± 20	-	-
2006 Mar 24.96	34.81	-	-	15 ± 18	-	-
2006 Mar 26.85	36.70	69 ± 69	5 ± 37	-	-	-
2006 Mar 31.89	41.74	-	-	22 ± 22	-	-
2006 Apr 9.84	50.70	-	-	25 ± 25	-	-
2006 Jun 2.67	104.52	-	-	17 ± 21	-	-

We used the standard continuum mode with 2×50 -MHz bands (VLA) and 350-MHz bandwidth (Ryle). At 22.5 GHz we used referenced pointing scans to correct for the systematic 10–20-arcsec pointing errors of the VLA antennas. We used the extra-galactic sources 3C 48 (J0137+331) and 3C 147 (J0542+498) for flux calibration, while the phase was monitored using J0319+190 (VLA) and J0326+1521 (Ryle). The data were reduced and analysed using the Astronomical Image Processing System. The flux density and uncertainty were measured from the resulting maps by fitting a gaussian model to the afterglow emission.

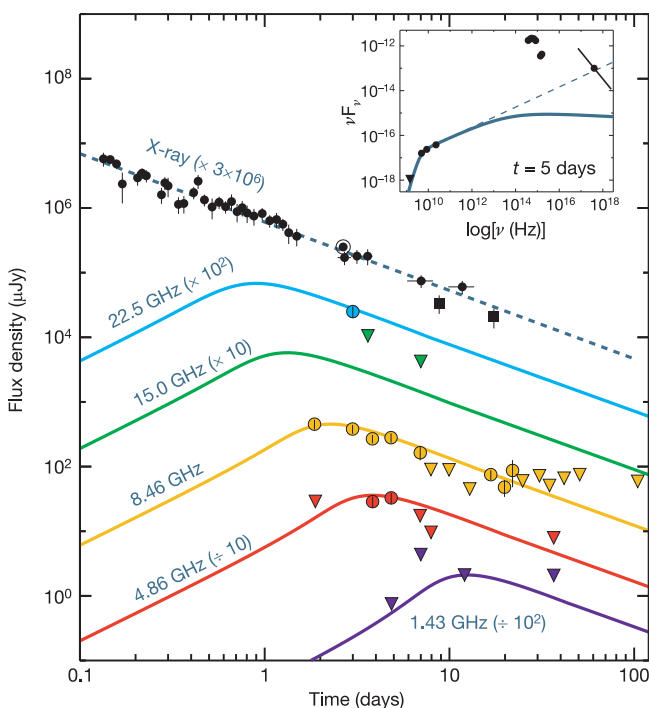
observations at 104 days show no evidence for a late-time increase in the radio flux, thus constraining the presence of additional ejecta components (off-axis jets; Fig. 2) spreading into our line-of-sight.

As can be seen from Fig. 1, the above synchrotron model is unable to explain the strong X-ray emission. Attributing the emission to scattering of supernova optical photons by the mildly relativistic ejecta requires an optical depth of $\tau \approx 10^{-4}$, too large to be produced by the shocked electrons which provide $\tau = nr\sigma_T \approx 10^{-7}$; here σ_T is the Thomson cross-section. We must therefore seek an entirely different origin for the observed X-rays.

At day 1, the steep X-ray spectrum roughly connects to the peculiar optical/ultraviolet component ($\beta_{\text{OX}} \approx -2$) observed⁵ to peak on this timescale. A similar steep near-infrared spectrum was seen¹⁵ in GRB 031203 at $t = 0.4$ days. Given that both GRB 031203 and XRF 060218 are sub-energetic events^{11,10}, we suggest that this mysterious steep

component is ubiquitous among sub-energetic GRBs and speculate that a central engine is the origin of this intense, long-lived emission. One particularly attractive possibility is a rapidly rotating (period P), highly magnetized (field strength B) neutron star: a magnetar. The spin-down power, $\dot{E} = 10^{45}(P/10)^{-4}(B/10^{15})^2 \text{ erg s}^{-1}$, where P is in milliseconds and B is in gauss, can explain the peculiar optical to X-ray integrated luminosity at 1 day, while the temporal evolution requires a braking index lower than three (magnetic dipole). We note that similarly low braking indices are measured for young Galactic pulsars (ref. 16 and references therein).

Combining the sky coverage and detection thresholds of γ -ray missions, we estimate the following sensitivity to the two exemplars

**Figure 1 | Radio and X-ray light-curves of XRF 060218.**

Radio measurements are summarized in Table 1. Upper limits are given as 3σ (inverted triangles). Error bars are 1σ . Solid lines are models of synchrotron emission from a spherical shock expanding into a wind-blown circumstellar medium ($n \propto r^{-2}$). At $t = 5$ days the radio spectrum peaks near 4 GHz owing to the synchrotron self-absorption frequency, ν_a . We assume that the energy density is partitioned between the relativistic electrons (energy distribution $N(\gamma) \propto \gamma^{-p}$ with $p \approx 2.1$) and magnetic field as $\epsilon_e = \epsilon_B = 0.1$. We find that $E_K \approx 2 \times 10^{48} \text{ erg}$ is coupled to ejecta with $\Gamma \approx 2.3$. The expansion, $r \propto t^m$, appropriate²⁵ for a core-collapse supernova explosion with a distribution of ejecta velocities, is fitted with $m \approx 0.85$. We infer a progenitor mass loss rate of $2 \times 10^{-7} M_\odot \text{ yr}^{-1}$ (for wind velocity, $v_w \approx 10^3 \text{ km s}^{-1}$). These parameters constrain the characteristic synchrotron frequency, $\nu_m \approx 0.3 \text{ GHz}$, and the synchrotron cooling frequency $\nu_c \approx 10^{14} \text{ Hz}$, at $t = 5$ days and thus $\nu_m < \nu_a$; consistent with the observed radio spectrum (inset, solid grey curve). A nearly identical fit is obtained for a trans-relativistic GRB blastwave expanding into a constant-density circumstellar medium²⁶ for parameters: $E_K \approx 1.2 \times 10^{48} \text{ erg}$, $n = 10^2 \text{ cm}^{-3}$, $\epsilon_e = \epsilon_B = 0.1$ and $p = 2.1$; in this case the mildly relativistic ejecta is assumed to expand with a single bulk Lorentz factor. These values constrain the geometry of the ejecta to be effectively spherical, $\theta_j \gtrsim 1.4$. The X-ray flux (XRT, circles; XMM, circled dot, scaled to XRT spectral model; CXO, squares) is significantly brighter than an extrapolation of the above model, as shown by the unusually flat radio to X-ray spectral index, $\beta_{\text{RX}} \approx -0.5$ (inset, grey dashed line), and the steep X-ray spectrum $\beta_X \approx -2.2$ (inset, black line), instead of $\beta_X \approx -1.1$ for typical GRBs. We suggest that the integrated optical to X-ray luminosity ($10^{44} \text{ erg s}^{-1}$; $2\text{--}10^4 \text{ eV}$) can be attributed to the spin-down power of a magnetar. By day 5, the optical/ultraviolet spectrum is dominated by the thermal supernova emission (inset).

of low-energy events (GRB 980425 and XRF 060218): 3.8×10^{-3} (BeppoSAX), 1.2×10^{-3} (HETE-2) and $3.7 \times 10^{-3} \text{ Gpc}^3 \text{ yr}^{-1}$ (Swift). Thus, the true rate of sub-energetic GRBs is $230^{+490}_{-190} \text{ Gpc}^3 \text{ yr}^{-1}$ (90% confidence range; see Supplementary Information I), about ten times more abundant than typical bright GRBs¹⁷, for which we use a mean inverse beaming factor of $\langle f_b^{-1} \rangle \sim 100$; here $f_b \equiv 1 - \cos \theta_j$. Separately, we note that sub-energetic GRBs could not be strongly beamed or the true rate of such events would exceed the local ($d \approx 100 \text{ Mpc}$) rate^{18,19} of type Ibc supernovae, $9^{+3}_{-5} \times 10^3 \text{ Gpc}^3 \text{ yr}^{-1}$.

Spectroscopy of the nearest GRB-associated supernovae (SN

1998bw, SN 2003dh, SN 2003lw and now SN 2006aj) reveals (see ref. 4 and references therein) remarkably broad absorption lines (indicative of fast ejecta) and may suggest that all GRB-supernovae are broad-lined. Locally, broad-lined events comprise²⁰ 5% of type Ibc supernovae. Thus, the rate of broad-lined events and sub-energetic GRBs are comparable, suggesting that all broad-lined supernovae harbour a long-lived central engine.

Radio observations of an extensive sample of 144 optically selected local type Ibc supernovae, however, suggest^{21,12} a different picture (Fig. 2). Not a single supernova (broad-lined or otherwise) shows strong early radio emission comparable to that seen in SN 1998bw and SN 2006aj. Thus, we constrain the volumetric rate of such events to be $\lesssim 300 \text{ Gpc}^3 \text{ yr}^{-1}$ (see Supplementary Information II), consistent with the rate of sub-energetic GRBs inferred above. Focusing on the broad-lined supernovae, less than one in three are similar to GRBs, indicating that broad lines cannot be used as a reliable proxy for a central engine.

The commonality between the three nearest events (GRB 980425, XRF 060218, GRB 031203) is their substantial ($E_K \gtrsim 10^{48} \text{ erg}$) mildly relativistic ($\Gamma \gtrsim 2$) ejecta and a smooth pulse profile for the prompt emission. These two clues lead us to suggest that the primary physical distinction between GRBs or XRFs and ordinary supernovae is the velocity profile of the ejecta. For the latter, hydrodynamic collapse

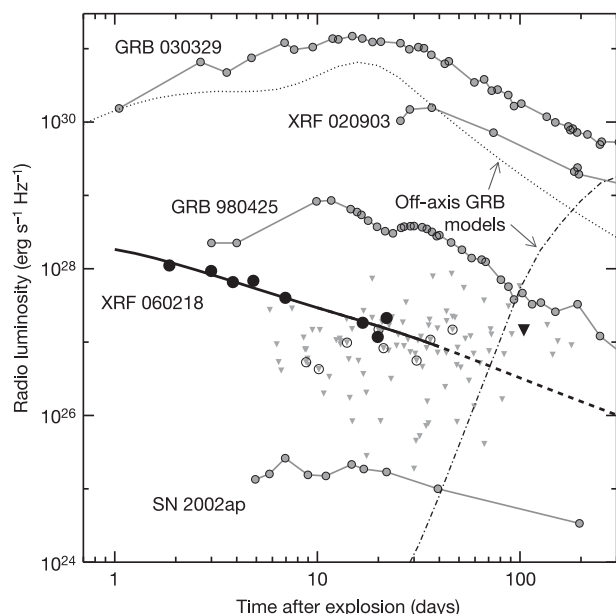


Figure 2 | Radio observations for a large sample of local type Ibc supernovae. Since 1999 we have been monitoring the radio emission from optically selected type Ibc supernovae with the VLA. We use radio luminosity as a proxy for mildly relativistic ejecta to quantify the fraction of type Ibc supernovae powered by central engines. Our observations of 144 supernovae show that most type Ibc supernovae do not produce strong radio emission and therefore show no evidence for a central engine. For comparison, we include the radio afterglows for nearby ($z \lesssim 0.25$) GRB 980425 and GRB 030329, and XRF 020903, all three of which show^{2,8,27} evidence for an engine-driven explosion. XRF 060218 is intermediate between GRBs and the broad-lined SN 2002ap, demonstrating that broad lines are not a reliable proxy for strong radio emission. Radio limits for other local broad-lined supernovae (encircled triangles) show that less than one in three of these events may have a radio luminosity comparable to XRF 060218 or GRB 980425 (90% confidence level). In addition, we show two 8.5-GHz model light curves for a typical GRB viewed away from the collimation axis. Both models adopt typical GRB parameters (see ref. 12 and references therein) of $\Gamma = 100$, $E_{K,iso} = 10^{53} \text{ erg}$, $n = 1 \text{ cm}^{-3}$, $\varepsilon_e = \varepsilon_B = 0.1$ and $p = 2.1$. In the first model we assume that the observed γ -ray and radio emission are produced by a GRB jet viewed from an angle $\theta_{obs} = 2\theta_j$; here θ_{obs} is the angle between our line-of-sight and the jet axis. In this scenario, the observed prompt emission properties (Δt , E_p , $E_{\gamma,iso}$) are related to the intrinsic values through the quantity $D \equiv [(\Gamma\theta_{obs} - \theta_j)]^{-2}$. For $D \approx 0.02$, the intrinsic properties for XRF 060218 would be typical for GRBs: $\Delta t \approx 40 \text{ s}$, $E_p \approx 250 \text{ keV}$, $E_{\gamma,iso} \approx 10^{53} \text{ erg}$, and $\theta_j \approx 4^\circ$. The resulting off-axis model (dotted line) is a factor of 10^3 brighter than the observed XRF 060218 radio light-curve and can therefore be ruled out. In the second model, we assume that in addition to the quasi-spherical mildly relativistic ejecta component producing the observed radio emission, XRF 060218 also harbours a strongly collimated relativistic jet directed significantly away from our line-of-sight. In this scenario, we expect^{13,12} a late-time radio re-brightening as the jet becomes non-relativistic and spreads sideways into our line-of-sight. Adopting $\theta_j = 4^\circ$ we find that our latest radio limit (104 days; black triangle) rules out an off-axis GRB with $\theta_{obs} \lesssim 60^\circ$ (dash-dotted line). We conclude that the XRF 060218 ejecta was quasi-spherical and intrinsically sub-energetic.

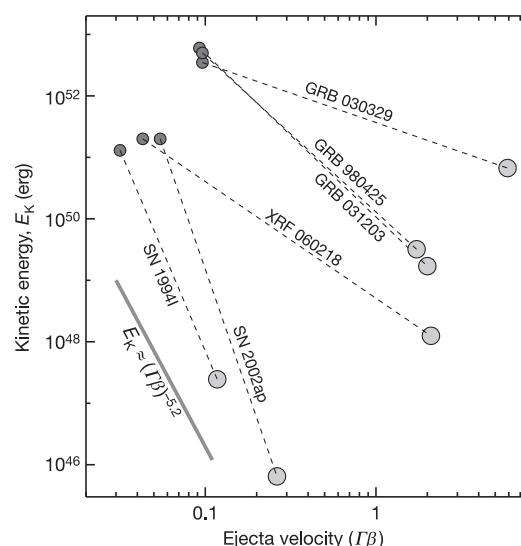


Figure 3 | Energy as a function of velocity for GRBs, XRFs, and type Ibc supernovae. Optical data (small dark circles) probe (see refs 28, 29 and references therein) the slowest ejecta in supernova explosions, which typically carry the bulk of the kinetic energy ($E_K = 0.3 M_{ej} v_{ej}^2 \approx 10^{51} \text{ erg}$). On the other hand, radio observations (large light circles) trace^{21,11,21,25,27} only the fastest ejecta in the explosion. For GRB 030329 and GRB 031203, $\Gamma \propto t^{-3/8}$; we adopt the bulk velocity of the relativistic ejecta at day 1 as inferred from radio modelling. For GRB 980425, XRF 060218, SN 2002ap and SN 1994I the bulk velocity is roughly constant on the timescale probed by the radio observations; we adopt the velocity at the radio peak time. Standard hydrodynamic collapse results³⁰ in a kinetic energy profile, $E_K \propto (\Gamma\beta)^{-5.2}$ (grey line), and thus a negligible fraction of the kinetic energy may be coupled to mildly relativistic ejecta, consistent with the radio observations of local type Ibc supernovae 1994I and 2002ap. In the case of typical GRBs (such as GRB 030329), however, the kinetic energy of the mildly relativistic ejecta is nearly comparable to that of the slower material, indicating the presence of a central engine. Because the origin of the relativistic flow is separate from the supernova, there is probably not a continuous distribution of matter between the two data points but rather distinct ejecta components. Sub-energetic bursts such as XRF 060218 are intermediate between these two classes and may indicate that their central engines are different than those of typical GRBs. We conclude that the minimum criteria for producing GRBs and XRFs is a mildly relativistic ($\Gamma \gtrsim 2$), quasi-spherical ejecta carrying at least 10^{48} erg .

requires that the ejecta energy is concentrated at low velocities, $E_K \propto (I\beta)^{-5.2}$. In comparison, the shallow velocity profiles inferred for GRBs and XRFs indicate that some other agent (an engine) enables coupling of copious energy to the relativistic material (Fig. 3).

We conclude by noting that magnetars constitute²² about 10% of the Galactic neutron-star birth-rate, and thus a similar fraction of type Ibc supernovae. This rate is similar to that of the sub-energetic GRBs. Furthermore, magnetars produce long-lived emission (see ref. 23 and references therein) and have been suggested²⁴ previously as candidate GRB progenitors. We therefore speculate that a magnetar central engine is what distinguishes sub-energetic GRBs from the cosmological bursts, which are thought to be powered by a black hole.

Received 18 April; accepted 13 July 2006.

- Galama, T. J. *et al.* An unusual supernova in the error box of the gamma-ray burst of 25 April 1998. *Nature* **395**, 670–672 (1998).
- Kulkarni, S. R. *et al.* Radio emission from the unusual supernova 1998bw and its association with the gamma-ray burst of 25 April 1998. *Nature* **395**, 663–669 (1998).
- Matheson, T. *et al.* Photometry and spectroscopy of GRB 030329 and its associated supernova 2003dh: the first two months. *Astrophys. J.* **599**, 394–407 (2003).
- Pian, E. *et al.* An optical supernova associated with the X-ray flash XRF 060218. *Nature* doi:10.1038/nature05082 (this issue).
- Campana, S. *et al.* The association of GRB 060218 with a supernova and the evolution of the shock wave. *Nature* doi:10.1038/nature04892 (this issue).
- Mirabal, N., Halpern, J. P., An, D., Thorstensen, J. R. & Terndrup, D. M. GRB 060218/SN 2006aj: a gamma-ray burst and prompt supernova at $z = 0.0335$. *Astrophys. J.* (submitted); preprint at (<http://www.arXiv.org/astro-ph/0603686>) (2006).
- Heise, J., in't Zand, J., Kippen, R. M. & Woods, P. M. in *Gamma-Ray Bursts in the Afterglow Era* (eds Costa, E., Frontera, F. & Hjorth, J.) 16–21 (Springer, Berlin/Heidelberg, 2001).
- Soderberg, A. M. *et al.* A redshift determination for XRF 020903: first spectroscopic observations of an X-ray flash. *Astrophys. J.* **606**, 994–999 (2004).
- De Luca, A. GRB 060218: analysis of the XMM-Newton observation. *GRB Circ. Netw.* **4853** (2006).
- Sazonov, S. Y., Lutovinov, A. A. & Sunyaev, R. A. An apparently normal γ -ray burst with an unusually low luminosity. *Nature* **430**, 646–648 (2004).
- Soderberg, A. M. *et al.* The sub-energetic γ -ray burst GRB 031203 as a cosmic analogue to the nearby GRB 980425. *Nature* **430**, 648–650 (2004).
- Soderberg, A. M., Nakar, E., Berger, E. & Kulkarni, S. R. Late-time radio observations of 68 type Ibc supernovae: strong constraints on off-axis gamma-ray bursts. *Astrophys. J.* **638**, 930–937 (2006).
- Waxman, E. The nature of GRB 980425 and the search for off-axis gamma-ray burst signatures in nearby type Ib/c supernova emission. *Astrophys. J.* **602**, 886–891 (2004).
- Sari, R., Piran, T. & Halpern, J. P. Jets in gamma-ray bursts. *Astrophys. J.* **519**, L17–L20 (1999).
- Malesani, D. *et al.* SN 2003lw and GRB 031203: A bright supernova for a faint gamma-ray burst. *Astrophys. J.* **609**, L5–L8 (2004).
- Lyne, A. G., Pritchard, R. S., Graham-Smith, F. & Camilo, F. Very low braking index for the VELA pulsar. *Nature* **381**, 497–498 (1996).
- Schmidt, M. Luminosity function of gamma-ray bursts derived without benefit of redshifts. *Astrophys. J.* **552**, 36–41 (2001).
- Cappellaro, E., Evans, R. & Turatto, M. A new determination of supernova rates and a comparison with indicators for galactic star formation. *Astron. Astrophys.* **351**, 459–466 (1999).
- Dahlen, T. *et al.* High-redshift supernova rates. *Astrophys. J.* **613**, 189–199 (2004).
- Podsiadlowski, P., Mazzali, P. A., Nomoto, K., Lazzati, D. & Cappellaro, E. The rates of hypernovae and gamma-ray bursts: implications for their progenitors. *Astrophys. J.* **607**, L17–L20 (2004).
- Berger, E., Kulkarni, S. R., Frail, D. A. & Soderberg, A. M. A radio survey of type Ib and Ic supernovae: Searching for engine-driven supernovae. *Astrophys. J.* **599**, 408–418 (2003).
- Gaensler, B. M. *et al.* A stellar wind bubble coincident with the anomalous X-ray pulsar 1E 1048.1-5937: are magnetars formed from massive progenitors? *Astrophys. J.* **620**, L95–L98 (2005).
- Hurley, K. *et al.* An exceptionally bright flare from SGR 1806-20 and the origins of short-duration γ -ray bursts. *Nature* **434**, 1098–1103 (2005).
- Usov, V. V. Millisecond pulsars with extremely strong magnetic fields as a cosmological source of gamma-ray bursts. *Nature* **357**, 472–474 (1992).
- Chevalier, R. A. Synchrotron self-absorption in radio supernovae. *Astrophys. J.* **499**, 810–819 (1998).
- Granot, J. & Sari, R. The shape of spectral breaks in gamma-ray burst afterglows. *Astrophys. J.* **568**, 820–829 (2002).
- Berger, E. *et al.* A common origin for cosmic explosions inferred from calorimetry of GRB030329. *Nature* **426**, 154–157 (2003).
- Baron, E., Branch, D., Hauschildt, P. H., Filippenko, A. V. & Kirshner, R. P. Spectral models of the type Ic supernova SN 1994I in M51. *Astrophys. J.* **527**, 739–745 (1999).
- Mazzali, P. A. *et al.* A neutron-star-driven X-ray Flash associated with supernova SN 2006aj. *Nature* doi:10.1038/nature05081 (this issue).
- Tan, J. C., Matzner, C. D. & McKee, C. F. Trans-relativistic blast waves in supernovae as gamma-ray burst progenitors. *Astrophys. J.* **551**, 946–972 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements GRB research at Caltech is supported in part by funds from NSF and NASA. We are, as always, indebted to S. Barthelmy and the GCN. The VLA is operated by the National Radio Astronomy Observatory, a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. A.M.S. and S.B.C. are supported by NASA Graduate Research Fellowships. E.B. and A.G.-Y. acknowledge support by NASA through a Hubble Fellowship grant. D.N.B. and J.A.N. acknowledge support by NASA.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.M.S. (ams@astro.caltech.edu).

LETTERS

A neutron-star-driven X-ray flash associated with supernova SN 2006aj

Paolo A. Mazzali^{1,3,4,5,6}, Jinsong Deng^{2,3,4,6}, Ken'ichi Nomoto^{3,4,6}, Daniel N. Sauer^{5,6}, Elena Pian^{5,6}, Nozomu Tominaga^{3,6}, Masaomi Tanaka³, Keiichi Maeda^{6,7} & Alexei V. Filippenko⁸

Supernovae connected with long-duration γ -ray bursts^{1–3} (GRBs) are hyper-energetic explosions resulting from the collapse of very massive stars ($\sim 40 M_{\odot}$, where $M_{\odot}$ is the mass of the Sun) stripped of their outer hydrogen and helium envelopes^{4–7}. A very massive progenitor, collapsing to a black hole, was thought to be a requirement for the launch of a GRB⁸. Here we report the results of modelling the spectra and light curve of SN 2006aj (ref. 9), which demonstrate that the supernova had a much smaller explosion energy and ejected much less mass than the other GRB-supernovae, suggesting that it was produced by a star whose initial mass was only $\sim 20 M_{\odot}$. A star of this mass is expected to form a neutron star rather than a black hole when its core collapses. The smaller explosion energy of SN 2006aj is matched by the weakness and softness¹⁰ of GRB 060218 (an X-ray flash), and the weakness of the radio flux of the supernova¹¹. Our results indicate that the supernova–GRB connection extends to a much broader range of stellar masses than previously thought, possibly involving different physical mechanisms: a ‘collapsar’ (ref. 8) for the more massive stars collapsing to a black hole, and magnetic activity of the nascent neutron star¹² for the less massive stars.

Like all other GRB-supernovae, SN 2006aj is of type Ic (ref. 9). Its spectra resemble those of the dim, broad-lined, non-GRB supernova SN 2002ap (refs 13, 14). However, SN 2006aj shows surprisingly weak oxygen lines for a type Ic supernova. For a comparison of the spectrum of SN 2006aj and those of SN 2002ap and of the GRB-supernova SN 1998bw, see Supplementary Information.

To reproduce the spectrum of SN 2006aj (ref. 9) we started from the model that was used for SN 2002ap (ref. 13), but to improve the spectral fits we reduced the masses of both oxygen and calcium significantly, and decreased the ejected mass M_{ej} and the kinetic energy E_K accordingly. The series of synthetic spectra is shown in Fig. 1.

A lack of oxygen lines in the spectrum suggests a small M_{ej} , but it does not necessarily mean absence of oxygen in the ejecta. Our model contains $\sim 1.3 M_{\odot}$ of oxygen. Oxygen is therefore still the dominant element, but its abundance relative to other (heavier) elements is much lower than in SN 2002ap or in the other GRB-supernovae. Modelling also indicates that oxygen is confined to high velocities (Fig. 1). A shell of oxygen comprising $\sim 0.1 M_{\odot}$ and expanding at velocities between 20,000 and 30,000 km s^{−1} is detected, which may be the result of the episode of interaction that was responsible for the early ultraviolet brightening¹⁰.

The spectroscopic results are confirmed by models of the light curve. A synthetic light curve computed using the one-dimensional density and chemical abundance structure obtained from the spectral

analysis reproduces the optical-infrared bolometric light curve of SN 2006aj (Fig. 2). For SN 2006aj we derive $M_{\text{ej}} \approx 2 M_{\odot}$ and $E_K \approx 2 \times 10^{51}$ erg. These values are much smaller than those of the other GRB-supernovae, which typically have $M_{\text{ej}} \approx 10 M_{\odot}$ and $E_K \approx 3 \times 10^{52}$ erg (refs 4–7). The smaller E_K and M_{ej} involved for SN 2006aj explain why the light curve evolves more rapidly than that of SN 2002ap: the timescale of the light curve depends in fact roughly on M_{ej}^3/E_K (ref. 15). The supernova ejecta contain $0.21 M_{\odot}$ of ^{56}Ni , which is responsible for the supernova luminosity. About $0.02 M_{\odot}$ of this is located above 20,000 km s^{−1} and causes the fast rise of the light curve. The presence of ^{56}Ni at high velocities is unlikely to be the result of a spherically symmetric explosion. In a realistic aspherical explosion, high-velocity ^{56}Ni may be copiously produced near the direction of the GRB jets¹⁶.

Observations in the nebular phase, when the forbidden [O I] 6,300 Å and 6,363 Å lines should be strong in emission, will be needed to determine more accurately the value of M_{ej} . Such observations, to be performed starting August 2006, will also be useful in studying any possible asymmetry and the orientation of the supernova with respect to the line of sight to the Earth, and thus to link the supernova with the GRB^{16,17}.

The properties of both the supernova (small energy, small ejected mass, low oxygen content) and of the GRB (unusually soft and long) seem to suggest that the GRB 060218–SN 2006aj event was not the same type of event as the other GRB-supernovae known thus far. The radio properties of SN 2006aj were also intermediate between those of the GRB-supernovae and of SN 2002ap (ref. 11).

One possibility is that the initial mass of the progenitor star was significantly smaller than in the other GRB-supernovae, and that the collapse/explosion generated less energy. A star with zero-age main-sequence mass of ~ 20 – $25 M_{\odot}$ would be at the boundary between collapse to a black hole or to a neutron star¹⁸. If the star collapsed only to a neutron star, more core material would be available to synthesize ^{56}Ni . For example, a star with initially $\sim 20 M_{\odot}$ would develop a carbon–oxygen core of $\sim 3.3 M_{\odot}$ (ref. 18). If core collapse left behind a neutron star of $\sim 1.4 M_{\odot}$, $\sim 1.3 M_{\odot}$ of oxygen and $\sim 0.6 M_{\odot}$ of heavier elements (including both intermediate-mass elements such as Si and Fe-group elements) could be ejected in the supernova, consistent with our results. Such a collapse is thought to give rise to an explosion of $E_K \approx 10^{51}$ erg (ref. 19), but there are indications of a spread in both E_K and the mass of ^{56}Ni synthesized²⁰. Additionally, magnetar-type activity may have been present, increasing the explosion energy¹². Magnetic activity may also have caused the very long duration of the γ -ray emission¹² and the mixing-out of ^{56}Ni required by the rapid rise of the light curve. It is also possible

¹Max-Planck Institut für Astrophysik, Karl-Schwarzschild Strasse 1, D-85748 Garching, Germany. ²National Astronomical Observatories, CAS, 20A Datun Road, Chaoyang District, Beijing 100012, China. ³Department of Astronomy, School of Science, and ⁴Research Center for the Early Universe, School of Science, University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan. ⁵Istituto Nazionale di Astrofisica-OATs, Via Tiepolo 11, I-34131 Trieste, Italy. ⁶Kavli Institute for Theoretical Physics, University of California, Santa Barbara, California 93106-4030, USA. ⁷Department of Earth Science and Astronomy, College of Arts and Sciences, University of Tokyo, Komaba 3-8-1, Meguro-ku, Tokyo 153-8902, Japan. ⁸Department of Astronomy, University of California, Berkeley, California 94720-3411, USA.

that in this weaker explosion the fraction of energy channelled to relativistic ejecta was smaller than in the classical GRB-supernova, giving rise to an X-ray flash (XRF)¹¹.

Another case of a supernova associated with an XRF has been reported²¹. The putative supernova, although poorly observed, was also most consistent with the properties of SN 2002ap (ref. 22). This may suggest that XRFs are associated with less-massive progenitor stars than those of canonical GRBs, and that the two groups may be differentiated by the formation of a magnetar²³ or a black hole, respectively. The properties of both the GRB and the supernova may scale with the mass of the progenitor²⁴. Still, the progenitor of SN

2006aj had been thoroughly stripped of its H and He envelopes. This is a general property of all GRB-supernovae known so far, and possibly a requirement for the emission of a high-energy transient, which may be more easily achieved in a binary system^{13,25,26}.

If the star was initially more massive ($\geq 25 M_{\odot}$), and it collapsed directly to a black hole as in the more powerful GRB-supernova events, a number of questions arise. Why was the energy of the explosion so small? Where did the large core mass end up? Continuing accretion onto the black hole could explain the missing mass. This might occur if the angular momentum of the core was smaller than in the more energetic cases. Other more exotic scenarios, such as merger models, might also work.

A case of a progenitor mass just exceeding the black-hole limit may be that of SN 2002ap. This supernova may not have produced a magnetar and an XRF, because it did not collapse to a neutron star but rather to a black hole¹³, yet at the same time the energies involved in the collapse may have been too small to give rise to a GRB.

In our scenario, some soft γ -ray repeaters energized by a magnetar^{12,27} may be remnants of GRB 060218-like events. Magnetars could thus generate a GRB at different times. As they are born, when they

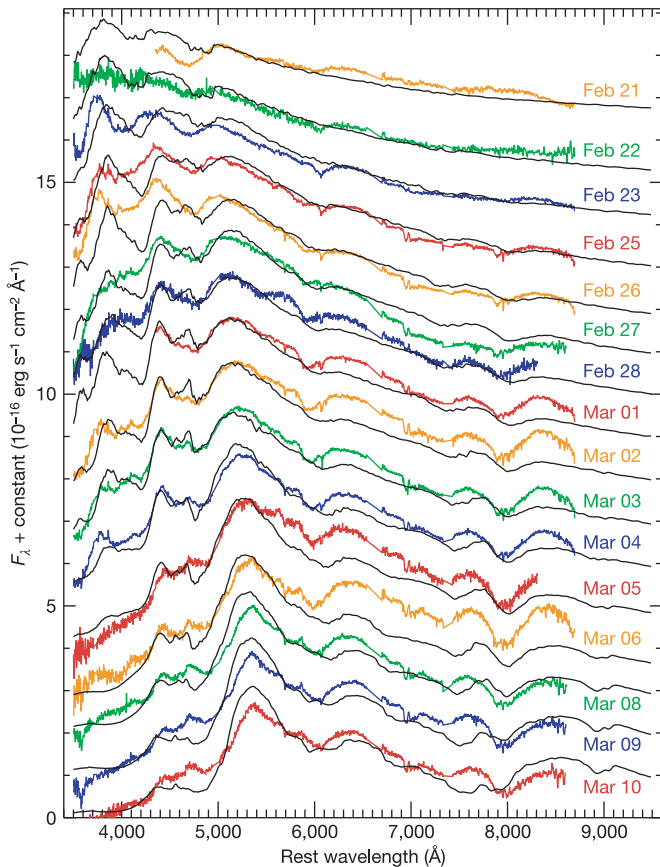


Figure 1 | Spectra of SN 2006aj and synthetic fits. The observed spectra of SN 2006aj (coloured traces) are calibrated in the V band, but elsewhere they may be distorted⁹, hence the poorer agreement in some of the red parts. Also, the blue part is not reliable shortward of $\sim 4,200 \text{ \AA}$. The synthetic spectra (black traces) were computed using our Monte Carlo spectrum synthesis code³⁰. Because of the spectroscopic and photometric similarity to SN 2002ap (ref. 14), we used a similar model of the explosion¹³, but to improve the match we reduced the masses of both oxygen and calcium significantly, and decreased M_{ej} and E_K accordingly. Our model has $M_{\text{ej}} \approx 2 M_{\odot}$ and $E_K \approx 2 \times 10^{51} \text{ erg}$. The strongest features in the spectra are due to lines of Fe II, Ti II, and in the later phases Ca II ($< 4,500 \text{ \AA}$), Fe III and Fe II (near $5,000 \text{ \AA}$), Si II (near $6,000 \text{ \AA}$), O I (near $7,500 \text{ \AA}$), and Ca II (near $8,000 \text{ \AA}$). The O I and Ca II lines become stronger at more advanced epochs, and are conspicuous because they form at a roughly constant wavelength, corresponding to a velocity ($\sim 25,000 \text{ km s}^{-1}$) higher than that of other lines. This indicates the presence of a shell of material, dominated by oxygen, at velocities between about $20,000$ and $25,000 \text{ km s}^{-1}$. This high-velocity material may result from the piling up of circumstellar material on the expanding ejecta. At modelled the spectrum by adding a small amount of mass ($\sim 0.10 M_{\odot}$) at $20,000 \leq v \leq 30,000 \text{ km s}^{-1}$. This results in an increased E_K ($\sim 2.5 \times 10^{51} \text{ erg}$). That the high-velocity material is mostly oxygen seems to confirm that both the outer supernova ejecta and the stellar wind were dominated by oxygen, and that the progenitor star was an early-type Wolf-Rayet star.

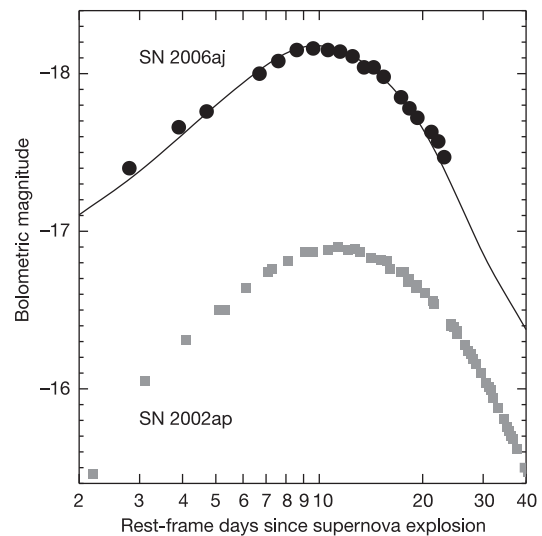


Figure 2 | The light curve of SN 2006aj. The bolometric light curve of SN 2006aj (circles) is compared with the model light curve (solid line), and with the bolometric light curve of SN 2002ap (squares). A supernova light curve is powered by γ -rays released in radioactive decays of freshly synthesized unstable ^{56}Ni to ^{56}Co and hence to stable ^{56}Fe . The γ -rays deposit in the dense ejecta, giving rise to a flux of optical photons. The light curve rises at first as the diffusion time of photons decreases, as the ejecta expand. A maximum is reached when the escaping photon luminosity approximately equals the deposited energy¹⁵. The light curve then declines as the density becomes low enough to allow significant γ -ray escape. The more massive the supernova ejecta and the smaller their kinetic energy, the more difficult it is for photons to escape, which means that the light curve reaches its maximum later and has a broader peak. The bolometric light curves were constructed by integrating the optical and near-infrared fluxes (for SN 2006aj, optical photometry obtained with the European Southern Observatory's (ESO) Very Large Telescope (VLT) and near-infrared photometry reported in the Gamma-Ray Burst Coordinates Network (GCN) were used), after correcting for the host-galaxy distance/redshift and the reddening towards the supernova—for SN 2006aj, 143 Mpc , $z = 0.0335$, and $E(B - V) = 0.13 \text{ mag}$ (ref. 9). The model light curve is synthesized from the one-dimensional density and chemical abundance structure of the best-fitting spectral models. It corresponds to $\sim 2 M_{\odot}$ ejecta expanding with a kinetic energy of $\sim 2 \times 10^{51} \text{ erg}$, having in total $\sim 0.2 M_{\odot}$ of ^{56}Ni . The small amount of mass and energy added by the inclusion of the outer oxygen shell (see Fig. 1) have a very limited impact on the light curve because the mass is located at low density and has low optical depth. The explosion of SN 2006aj is assumed to coincide in time with the GRB.

have a very short spin period (~ 1 ms), an XRF (or a soft GRB) may be produced as in SN 2006aj–GRB 060218. Later (after more than 1,000 years), when their spin rate is much lower, they could produce short-hard GRBs by a giant flare²⁸. Finally, if the progenitor star had a massive companion in a close binary system, as may be required for the outer envelope to be stripped and a long-duration GRB or XRF to be produced²⁶, the system may evolve to a close double-neutron-star system. When the two neutron stars finally merge, a short-hard GRB may again be produced²⁹.

Received 20 March; accepted 3 July 2006.

- Galama, T. J. *et al.* An unusual supernova in the error box of the gamma-ray burst of 25 April 1998. *Nature* **395**, 670–672 (1998).
- Stanek, K. Z. *et al.* Spectroscopic discovery of the supernova 2003dh associated with GRB 030329. *Astrophys. J.* **591**, L17–L20 (2003).
- Malesani, J. *et al.* SN 2003lw and GRB 031203: a bright supernova for a faint gamma-ray burst. *Astrophys. J.* **609**, L5–L8 (2004).
- Iwamoto, K. *et al.* A hypernova model for the supernova associated with the gamma-ray burst of 25 April 1998. *Nature* **395**, 672–674 (1998).
- Mazzali, P. A. *et al.* The type Ic supernova SN 2003dh/GRB 030329. *Astrophys. J.* **599**, L95–L98 (2003).
- Deng, J. *et al.* On the light curve and spectrum of SN 2003dh separated from the optical afterglow of GRB 030329. *Astrophys. J.* **624**, 898–905 (2005).
- Mazzali, P. A. *et al.* Models for the type Ic hypernova SN 2003lw associated with GRB 031203. *Astrophys. J.* **645**, 1323–1330 (2006).
- MacFadyen, A. E. & Woosley, S. E. Collapsars: gamma-ray bursts and explosions in 'Failed Supernovae'. *Astrophys. J.* **524**, 262–289 (1999).
- Pian, E. *et al.* An optical supernova associated with the X-ray flash XRF 060218. *Nature* doi:10.1038/nature05082 (this issue).
- Campana, S. *et al.* The association of GRB 060218 with a supernova and the evolution of the shock wave. *Nature* doi:10.1038/nature04892 (this issue).
- Soderberg, A. M. *et al.* Relativistic ejecta from X-ray flash XRF 060218 and the rate of cosmic explosions. *Nature* doi:10.1038/nature05087 (this issue).
- Thompson, T. A., Chang, P. & Quataert, E. Magnetar spin-down, hyperenergetic supernovae, and gamma-ray bursts. *Astrophys. J.* **611**, 380–393 (2004).
- Mazzali, P. A. *et al.* The type Ic hypernova SN 2002ap. *Astrophys. J.* **572**, L61–L64 (2002).
- Foley, R. J. *et al.* Optical photometry and spectroscopy of the SN 1998bw-like type Ic supernova 2002ap. *Publ. Astron. Soc. Pacif.* **115**, 1220–1235 (2003).
- Arnett, W. D. Type I supernovae. I—Analytic solutions for the early part of the light curve. *Astrophys. J.* **253**, 785–797 (1982).
- Maeda, K. *et al.* Explosive nucleosynthesis in aspherical hypernova explosions and late-time spectra of SN 1998bw. *Astrophys. J.* **565**, 405–412 (2002).
- Mazzali, P. A. *et al.* An asymmetric energetic type Ic supernova viewed off-axis, and a link to gamma ray bursts. *Science* **308**, 1284–1287 (2005).
- Tominaga, N. *et al.* The unique type Ib supernova 2005bf: a WN star explosion model for peculiar light curves and spectra. *Astrophys. J.* **633**, L97–L100 (2005).
- Nomoto, K. *et al.* A carbon–oxygen star as progenitor of the type-Ic supernova 1994I. *Nature* **371**, 227–229 (1994).
- Hamuy, M. Observed and physical properties of core-collapse supernovae. *Astrophys. J.* **582**, 905–914 (2003).
- Fynbo, J. P. U. *et al.* On the afterglow of the X-ray flash of 2003 July 23: photometric evidence for an off-axis gamma-ray burst with an associated supernova? *Astrophys. J.* **609**, 962–971 (2004).
- Tominaga, N. *et al.* Supernova light-curve models for the bump in the optical counterpart of X-ray flash 030723. *Astrophys. J.* **612**, L105–L108 (2004).
- Nakamura, T. A model for non high energy gamma ray bursts and sources of ultra high energy cosmic rays. *Prog. Theor. Phys.* **100**, 921–929 (1998).
- Nomoto, K. *et al.* Hypernovae and gamma-ray bursts. *Astrophys. Space Sci.* **298**, 81–86 (2005).
- Nomoto, K., Iwamoto, K. & Suzuki, T. The evolution and explosion of massive binary stars and Type Ib-Ic-IIb-III supernovae. *Phys. Rep.* **256**, 173–191 (1995).
- Fryer, C., Woosley, S. E. & Hartmann, D. H. Formation rates of black hole accretion disk gamma-ray bursts. *Astrophys. J.* **526**, 152–177 (1999).
- Thompson, C. & Duncan, R. C. The soft gamma repeaters as very strongly magnetized neutron stars—I. Radiative mechanism for outbursts. *Mon. Not. R. Astron. Soc.* **275**, 255–300 (1995).
- Hurley, K. *et al.* An exceptionally bright flare from SGR 1806-20 and the origins of short-duration γ -ray bursts. *Nature* **434**, 1098–1103 (2005).
- Narayan, R., Paczynski, B. & Piran, T. Gamma-ray bursts as the death throes of massive binary stars. *Astrophys. J.* **395**, L83–L86 (1992).
- Mazzali, P. A. Applications of an improved Monte Carlo code to the synthesis of early-time supernova spectra. *Astron. Astrophys.* **363**, 705–716 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Kulkarni, C. Fryer, T. Janka, W. Hillebrandt and C. Kouveliotou for many discussions. This work was supported in part by the European Union, by the JSPS and MEXT in Japan, and by the USA NSF.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.A.M. (mazzali@mpa-garching.mpg.de) or J.D. (jsdeng@bao.ac.cn).

Optical detection of liquid-state NMR

I. M. Savukov¹, S.-K. Lee¹ & M. V. Romalis¹

Nuclear magnetic resonance (NMR) in liquids and solids is primarily detected by recording the net dipolar magnetic field outside the spin-polarized sample. But the recorded bulk magnetic field itself provides only limited spatial or structural information about the sample. Most NMR applications rely therefore on more elaborate techniques such as magnetic field gradient encoding¹ or spin correlation spectroscopy², which enable spatially resolved imaging and molecular structure analysis, respectively. Here we demonstrate a fundamentally different and intrinsically information-rich modality of detecting NMR, based on the rotation of the polarization of a laser beam by the nuclear spins in a liquid sample. Optical NMR detection has in fact a long history in atomic vapours with narrow resonance lines^{3,4}, but has so far only been applied to highly specialized condensed matter systems such as quantum dots⁵. It has been predicted⁶ that laser illumination can shift NMR frequencies and thus aid detection, but the effect is very small and has never been observed. In contrast, our measurements on water and liquid ¹²⁹Xe show that the complementary effect—the rotation of light polarization by nuclear spins—is readily measurable, and that it is enhanced dramatically in samples containing heavy nuclei. This approach to optical NMR detection should allow correlated optical and NMR spectroscopy on complex molecules, and continuous two-dimensional imaging of nuclear magnetization with spatial resolution limited only by light diffraction.

The connection between light and liquid-state NMR was initially explored when it was suggested that NMR frequencies could be shifted by illumination with a circularly polarized laser far-detuned from optical resonances^{7,8}. However, more detailed experimental⁹ and theoretical^{10–14} work showed that these frequency shifts are extremely small, at most of the order of 10^{−5} Hz, and cannot be detected with present techniques. But the complementary magneto-optic effect, the rotation of far-off-resonance light polarization caused by nuclear spins, is readily measurable with a simple experimental apparatus because of the high density of nuclear spins in a liquid.

Both magneto-optic effects discussed can be related to the Faraday effect (the rotation of the plane of polarization of a light beam by a magnetic field). Nuclear magnetization in a liquid induces a magnetic field B_M that leads to optical rotation $\phi = lVB_M$, which is proportional to the Verdet constant V and the length of the sample l . Illumination by circularly polarized light induces electron spin magnetization in the excited state through the inverse Faraday effect¹⁵ that generates a magnetic field, causing NMR frequency shifts. The induced magnetic field B_M can be divided into a local contact field and a distant dipolar field. For example, for a long cylindrical sample with uniform magnetization M parallel to its axis, the classical magnetic field $B_M = 4\pi M$ consists of a contact interaction $B_c = 8\pi M/3$ and a distant dipolar field $B_d = 4\pi M/3$. The distant dipolar field depends on the shape of the sample (for example, it averages to zero for a spherical sample¹⁶), while the contact term can be enhanced or suppressed depending on the overlap of the wavefunction of the

virtual electron excitation created by the laser and the nuclear spin. We find that nuclear-spin optical rotation (NSOR) from ¹H in water is comparable to the size expected from the Faraday effect assuming no enhancement of the contact interaction, while for heavier atoms the contact term enhancement increases with the atomic number Z , making NSOR 135 times larger than Faraday rotation in liquid ¹²⁹Xe.

More formally, magneto-optic effects discussed in this Letter can be expressed in terms of the vector atomic polarizability α (ref. 17). The interaction energy for atoms in an oscillating electric field $E = (E_0/2)(\epsilon e^{-i\omega t} + \epsilon^* e^{i\omega t})$ is given by $H = -(E_0^2/4)\alpha \cdot s$, where s is the average photon spin, $s = i\epsilon \times \epsilon^*$, whereas rotation of light polarization is determined by the vector susceptibility of the medium, $\chi = N\alpha$, where N is the number density of atoms. For atoms with a nuclear spin I and a ¹S₀ electronic ground state, the vector polarizability as a function of laser frequency ω can be written as^{18,19}:

$$\alpha \equiv \alpha_v I = \frac{2\omega r_e c^2}{\hbar} \sum_k \frac{f_k a_k}{(\omega_k^2 - \omega^2)^2} I \quad (1)$$

where the sum is taken over dipole-transition-allowed excited states ($J = 1$, $L = 1$, $S = 0$) with resonance frequencies ω_k , oscillator strengths f_k and hyperfine coupling constants given by $H_k^{\text{hf}} = a_k I \cdot I$, and r_e is the classical electron radius. Hence, the NMR frequency shift for right circularly polarized light ($s = -1$) with intensity I_0 is given by:

$$\Delta\nu = \frac{I_0}{\hbar n c} \alpha_v \quad (2)$$

where n is the index of refraction of the liquid. The polarization rotation angle for a linearly polarized beam propagating in the z direction through the medium of length l is given by:

$$\phi = -\frac{2\pi\omega l N}{nc} \alpha_v \langle I_z \rangle \quad (3)$$

Thus, laser-induced NMR shift and NSOR depend on the same vector polarizability and can be easily related to each other. The polarizability is proportional to the strength of the hyperfine interaction and increases for heavier atoms. Our measurements in liquid ¹²⁹Xe and water illustrate the range of possible effects.

Spin-polarized ¹²⁹Xe was produced by spin-exchange with optically pumped Rb vapour²⁰, and collected as liquid in a 1 cm × 1 cm × 0.8 cm glass cell (Fig. 1a). Optical rotation was measured with a photoelastic modulation technique²¹ using lasers at three wavelengths (532 nm, 770 nm and 1,064 nm), with laser power ranging from 2 mW to 10 mW and beam size of the order of 1 mm. The laser beam was directed perpendicular to a static magnetic field $B_0 = 125$ mG and NSOR was detected at 147 Hz while ¹²⁹Xe spins were locked to an oscillating transverse magnetic field $B_1 = 7$ mG. The optical measurements were alternated with traditional NMR detection using a SQUID (superconducting quantum interference device) magnetometer for ¹²⁹Xe polarization calibration¹⁶. For optical measurements, the frequency of the oscillating transverse field was slowly swept to the NMR resonance at 147 Hz, left on resonance for 40 s, and swept back

¹Department of Physics, Princeton University, Princeton, New Jersey 08544, USA.

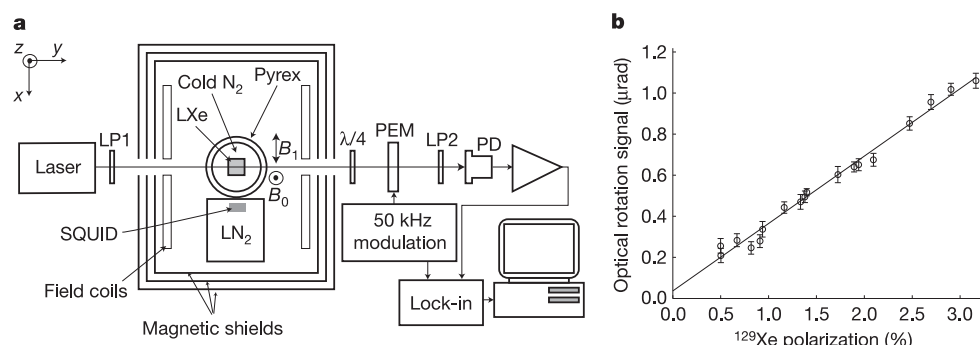


Figure 1 | Detection of nuclear-spin-induced optical rotation in liquid Xe.

a, Diagram showing the experimental set-up for the detection of nuclear-spin-induced optical rotation in liquid Xe (LXe). ^{129}Xe in natural abundance is polarized in a separate apparatus, introduced into a rectangular glass cell inside magnetic shields and maintained at -90°C by flowing cold N_2 gas through a double-wall Pyrex tube. Laser light is polarized by a linear polarizer (LP1) and the polarization rotation caused by liquid ^{129}Xe is converted to a change in intensity with a quarter-waveplate ($\lambda/4$), a photoelastic modulator (PEM) operating at 50 kHz and another linear polarizer (LP2). The light intensity is detected by a photodiode (PD) and

demodulated with a lock-in amplifier. Magnetic fields B_0 and B_1 are applied with field coils inside magnetic shields. A SQUID magnetometer operating in liquid nitrogen (LN_2) is used to independently measure Xe polarization. **b**, Data obtained with this set-up at 770 nm, and shown as the amplitude of the optical rotation signal as a function of ^{129}Xe polarization. The error bars correspond to ± 1 s.d. statistical uncertainty after averaging for 40 s with polarization rotation noise of $1 \times 10^{-7} \text{ rad Hz}^{-1/2}$ due to acoustical vibrations near the 147 Hz ^{129}Xe NMR frequency. The solid line is a linear fit with an intercept that is consistent with zero within errors.

off resonance. For calibration, a free induction decay signal was recorded following a 7° tip with a resonant radio-frequency pulse. Figure 1b shows the optical rotation angle as a function of ^{129}Xe polarization recorded during slow decay of nuclear polarization.

NSOR in water was detected using a different apparatus, shown in Fig. 2a. Water continuously flowed through a container placed inside a 9 T superconducting magnet to polarize ^1H spins, and then through a 50-cm-long glass tube held in a field $B_0 = 5 \text{ G}$. A transverse magnetic field $B_1 = 0.17 \text{ G}$ oscillating at 21 kHz was also continuously applied so that proton spins were adiabatically locked to the rotating field as they flowed into the apparatus. The degree of nuclear spin polarization along the path of the laser beam was independently measured using a non-resonant solenoidal NMR coil wound around the glass tube and connected to a high-input-impedance lock-in amplifier. Polarization loss during flow, inefficiency of adiabatic fast passage and broadening by magnetic field gradients resulted in a rotating nuclear spin polarization corresponding to a 1.5 T field. To avoid spurious cross-talk signals, the B_0 magnetic field was modulated on and off the proton NMR resonance at 8 Hz. The optical

rotation signal was measured using a balanced polarimeter with sensitivity limited by photon shot noise. Optical rotation and pick-up coil signals were recorded with a lock-in amplifier for several thousand seconds. Figure 2b shows the spectrum of the optical rotation signal with a peak at the 8 Hz modulation frequency. The water flow in the tube was in the turbulent regime (Reynolds number, 8,000); several measurements were made with different paths of the laser beam through the tube to verify that water had a uniform transverse distribution of polarization within 10% measurement error. No optical signal was detected when using circularly polarized light, thereby verifying absence of electronic cross-talk.

The optical rotation signals, normalized to 1 mol l^{-1} (1 M) concentration of fully polarized spins, are shown in Fig. 3 for ^{129}Xe in liquid xenon and ^1H in water, together with theoretical estimates. The size of the NSOR can be estimated from equation (1) if the oscillator strengths and hyperfine interaction constants are known. The L - S coupling scheme, used to derive equation (1), is not very accurate for ^{129}Xe owing to large relativistic effects, but can be expected to give a reasonable estimate. For example, the Verdet

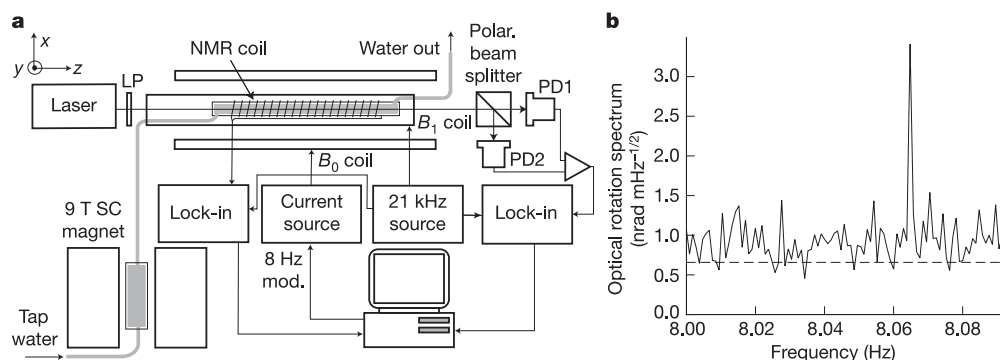


Figure 2 | Detection of nuclear-spin-induced optical rotation in water.

a, Diagram showing the experimental set-up for detection of optical rotation in water. Water is spin-polarized by flowing through a superconducting (SC) magnet and is adiabatically spin-locked to a field B_1 oscillating at 21 kHz in the y direction as it flows into a cylindrical glass tube placed in a uniform field B_0 in the x direction. The polarization rotation of laser light polarized with a linear polarizer (LP) is measured with a balanced polarimeter consisting of a polarizing beam splitter cube and two photodiodes (PD1 and PD2) and a lock-in amplifier referenced to the NMR frequency. The NMR

signal is also independently measured with a pick-up coil wound around the glass tube. A current source modulates the B_0 field on and off resonance at 8 Hz to avoid cross-talk. **b**, Data collected at 770 nm, and displayed as the Fourier spectral density of the optical rotation lock-in amplifier output for water with proton polarization $P = 5.3 \times 10^{-6}$. Taking into account the definite phase of the NMR signal, the signal-to-noise ratio after 1,000 s of averaging is equal to 4.5. The dashed line is the shot noise level for 2.9 mW detected laser power, calculated from the photocurrent in the photodetectors.

constant of Xe, calculated using the same equation with H^{hf} replaced by $H^{\text{B}} = \mu_{\text{B}} \mathbf{L} \cdot \mathbf{B}$, gives a result 40% smaller than the measured Verdet constant²², as could be expected, as contributions of higher excited states and the continuum are not included. However, in ^{129}Xe there is a substantial cancellation of contributions to the vector polarizability proportional to the nuclear spin between transitions to 6s and 5d excited states because they have similar oscillator strengths but opposite signs of the hyperfine constants, a_k . Using hyperfine constants²³ and oscillator strengths for a few of the lowest excited states²⁴, we estimate from equations (1) and (3) optical rotation of $3.5 \times 10^{-5} \text{ rad cm}^{-1}$ for a 1 M concentration of fully polarized ^{129}Xe spins at 770 nm, significantly larger than the measured value of $(5.8 \pm 0.6) \times 10^{-6} \text{ rad cm}^{-1} \text{ M}^{-1}$. This indicates that other atoms of similar nuclear charge but without cancellation between s and d states can have substantially larger optical rotation than ^{129}Xe .

The NMR frequency shift caused by circularly polarized light has been calculated for ^{129}Xe and other noble gases using several *ab initio* methods¹⁴. The results are converted to NSOR using equations (2) and (3), and plotted in Fig. 3a for two different sets of orbitals used in the multi-configuration self-consistent field (MCSCF) method. The excellent agreement with measured values may be somewhat fortuitous, as relativistic corrections are estimated¹⁴ to be of the order of 100%.

The size of NSOR in water cannot be easily estimated from equation (1) because the excited molecular level structure is very complicated and not all hyperfine constants are known. However, the measured optical rotation is in very good agreement with the size

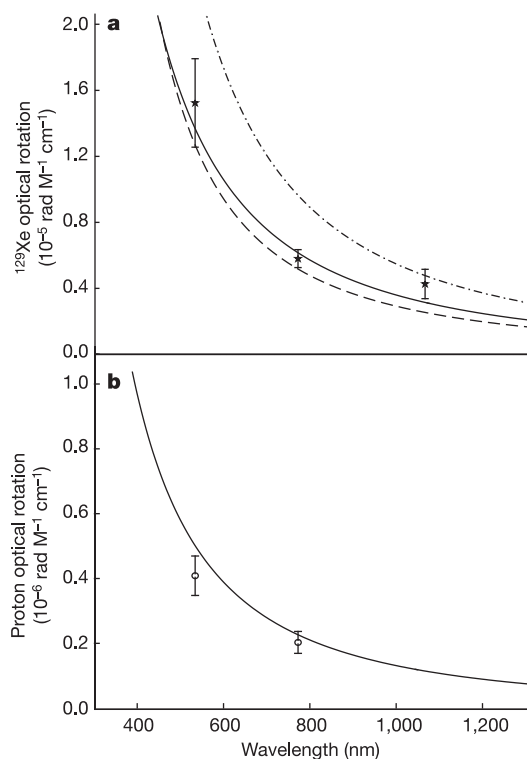


Figure 3 | Dependence of optical rotation on laser wavelength. Shown are the optical rotation angle per unit length for 1 M concentration of fully polarized ^{129}Xe spins in liquid Xe (a) and protons in water (b) as a function of laser wavelength. The error bars include combined statistical (± 1 s.d.) and systematic uncertainties. In a, the solid line is a fit to the wavelength dependence given by equation (1) with $\hbar\omega_k = 10$ eV. The results of a MCSCF *ab initio* calculation¹⁴ using complete active space 'CAS' and restricted active space 'RAS-III' of atomic orbitals are shown with dashed and dash-dotted lines, respectively. In b, the solid line is the optical rotation expected from nuclear magnetization due to the Faraday effect, assuming $B = 4\pi M$ in a long cylindrical cell, that is, $\kappa_{\text{H}} = 1$.

expected from the Faraday effect assuming $B = 4\pi M$, as shown with the solid line in Fig. 3b. This indicates that the excited-state electron wavefunction is not strongly enhanced at the location of the protons. In contrast, for ^{129}Xe the measured NSOR is more than 100 times larger than the rotation expected from the Faraday effect. The observed NSOR can be interpreted as an enhancement by a factor κ_{Xe} of the contact magnetic field $B_{\text{c}} = 8\pi\kappa_{\text{Xe}}M/3$, in analogy with the Fermi-contact interaction between alkali-metal atoms and noble gas nuclei²⁵. Our measurements give $\kappa_{\text{Xe}} = 135 \pm 13$, which is of the same order of magnitude as the enhancement for the alkali-metal electron wavefunctions during collision with ^{129}Xe atoms²⁵. Our measurements can also be compared with an *ab initio* calculation of laser-induced NMR frequency shifts in CS_2 (ref. 12). The calculated frequency shifts, converted to NSOR using equations (2) and (3) and compared with the expected Faraday rotation in CS_2 , give contact enhancement factors for ^{13}C and ^{33}S of $\kappa_{\text{C}} = 4.2$ and $\kappa_{\text{S}} = 14.7$, respectively, confirming the general trend of increase in κ with the nuclear charge.

Although in this first demonstration of NSOR the signal-to-noise ratio (SNR) is significantly lower than for traditional magnetic detection, it could be improved in several ways. Decreasing the laser wavelength (λ) increases the signal as $1/\lambda^2$ far from optical resonances, and even faster closer to resonances. At typical NMR frequencies, it should also be easy to realize photon shot-noise sensitivity for higher laser powers. The sensitivity of rotation measurements could be further improved by using a multi-pass or an optical cavity arrangement. For example, an effective optical path of 9 m has been demonstrated for simple organic liquids using a 1-cm-long cell placed in an optical cavity²⁶. The sample volume could be reduced by focusing the light into a thin capillary with a volume $V \approx 2\lambda^2 l$ determined by diffraction losses, where l is the path length. For a 1-cm-long capillary sample with a volume of 100 nl in an optical cavity with an effective length of 10 m, placed in a 10 T magnetic field and probed with 1 W of light at 400 nm wavelength, we estimate (based on scaling of our current results) an SNR of 1,000 for proton spins in water after 1 s of averaging; the SNR would be further enhanced for heavier nuclei. Another promising technique is to use a liquid-filled hollow optical fibre. Light guiding and detection of optical rotation in a liquid-filled photonic bandgap fibre has recently been demonstrated^{27,28}. The required sample volume for a single-mode hollow fibre is of the order of $V \approx 4\lambda^2 l$. For example, the same SNR of 1,000 could be achieved with a 10 nl sample volume using a 10-m-long fibre that can be coiled inside a magnet. For solutions containing molecules with a large molar mass, the maximum interaction length will be reduced owing to Rayleigh scattering, but the SNR per unit sample volume in a hollow fibre remains the same. Ultimate sensitivity with picolitre samples can be obtained by using a hollow fibre with mirrors at both ends²⁹. Thus, NSOR could be detected from small samples with a sensitivity higher than, or comparable to, that obtained in micro-coil NMR³⁰.

The optical rotation technique demonstrated here also has several unique advantages for the detection of NMR signals in transparent samples compared with magnetic detection. Using a two-dimensional photodiode array or CCD camera, one could obtain a real-time two-dimensional image of the nuclear magnetization without application of magnetic field gradients. The spatial resolution of such an image is in principle limited only by light diffraction. With a constant field gradient in the direction of the laser beam, one could also obtain a three-dimensional image of the precessing magnetization with high spatial resolution. For heavy atoms, the NSOR signal is significantly enhanced compared with the rotation due to the Faraday effect, allowing NMR signals to be detected in the presence of large magnetic field noise or radio-frequency fields at the NMR frequency. Another class of possible applications involves studies of correlation between optical and NMR spectroscopy. If the laser frequency is tuned near an optical resonance, the NSOR signal will be preferentially enhanced for nuclear spins that have a large overlap with the

excited-state electron wavefunction created by the optical excitation. Such two-dimensional optical-NMR spectroscopy could be useful for interpretation of circular dichroism data in complex molecules.

Received 15 April; accepted 11 July 2006.

- Lauterbur, P. C. Image formation by induced local interactions—examples employing nuclear magnetic resonance. *Nature* **242**, 190–191 (1973).
- Aue, W. P., Bartholdi, E. & Ernst, R. R. Two-dimensional spectroscopy—application to nuclear magnetic resonance. *J. Chem. Phys.* **64**, 2229–2246 (1976).
- Manuel, J. & Cohen-Tannoudji, C. Optical detection of magnetic resonance by modulation of paramagnetic Faraday effect transverse to Larmor frequency. *Compt. Rend.* **257**, 413–416 (1963).
- Romalis, M. V., Griffith, W. C., Jacobs, J. P. & Fortson, E. N. New limit on the permanent electric dipole moment of ^{199}Hg . *Phys. Rev. Lett.* **86**, 2505–2508 (2001).
- Kikkawa, J. M. & Awschalom, D. D. All-optical magnetic resonance in semiconductors. *Science* **287**, 473–476 (2000).
- Buckingham, A. D. & Parlett, L. C. High-resolution nuclear magnetic resonance spectroscopy in a circularly polarized laser beam. *Science* **264**, 1748–1750 (1994).
- Evans, M. W. Optical phase conjugation in nuclear magnetic resonance: Laser NMR spectroscopy. *J. Phys. Chem.* **95**, 2256–2260 (1991).
- Warren, W. S., Mayr, S., Goswami, D. & West, A. P. Jr Laser-enhanced NMR spectroscopy. *Science* **255**, 1683–1685 (1992).
- Warren, W. S., Goswami, D. & Mayr, S. Laser enhanced NMR spectroscopy, revisited. *Mol. Phys.* **93**, 371–375 (1998).
- Harris, R. A. & Tinoco, I. Laser-perturbed nuclear magnetic resonance spectroscopy and the conservation of parity. *J. Chem. Phys.* **101**, 9289–9294 (1994).
- Buckingham, A. D. & Parlett, L. C. The effect of circularly polarized light on NMR spectra. *Mol. Phys.* **91**, 805–813 (1997).
- Jaszuński, M. & Rizzo, A. A study of the effect of circularly polarized light on NMR spectra and related properties of CS_2 . *Mol. Phys.* **96**, 855–861 (1999).
- Li, L., He, T., Chen, D., Wang, X. & Liu, F.-C. Laser-induced NMR shift for Hg^{199} atom. *J. Phys. Chem. A* **102**, 10385–10390 (1998).
- Romero, R. H. & Vaara, J. Laser-induced splittings in the nuclear magnetic resonance spectra of the rare gas atoms. *Chem. Phys. Lett.* **400**, 226–230 (2004).
- Van der Ziel, P., Pershan, P. S. & Malmstrom, L. D. Optically-induced magnetization resulting from inverse Faraday effect. *Phys. Rev. Lett.* **15**, 190–193 (1965).
- Ledbetter, M. P. & Romalis, M. V. Nonlinear effects from dipolar interactions in hyperpolarized liquid ^{129}Xe . *Phys. Rev. Lett.* **89**, 287601 (2002).
- Happer, W. & Mathur, B. S. Effective operator formalism in optical pumping. *Phys. Rev.* **163**, 12–25 (1967).
- Buckingham, A. D. Permanent and induced molecular moments and long-range intermolecular forces. *Adv. Chem. Phys.* **12**, 107–142 (1967).
- Barron, L. D. *Molecular Light Scattering and Optical Activity* (Cambridge Univ. Press, New York, 1982).
- Driehuys, B. et al. High-volume production of laser-polarized ^{129}Xe . *Appl. Phys. Lett.* **69**, 1668–1670 (1996).
- Zavattini, E. et al. Experimental observation of optical rotation generated in vacuum by a magnetic field. *Phys. Rev. Lett.* **96**, 110406 (2006).
- Ingersoll, L. R. & Liebenberg, D. H. Faraday effect in gases and vapors. II. *J. Opt. Soc. Am.* **46**, 538–542 (1956).
- Luc-Koenig, E. A priori study of the influence of relativistic effects on the hyperfine structure of ^{129}Xe and ^{131}Xe . *J. Phys. E* **33**, 847–852 (1972).
- Yoshino, K. & Freeman, D. E. Absorption spectrum of xenon in the vacuum-ultraviolet region. *J. Opt. Soc. Am. B* **2**, 1268–1274 (1985).
- Schaefer, S. R. et al. Frequency shifts of the magnetic-resonance spectrum of mixtures of nuclear spin-polarized noble gases and vapors of spin-polarized alkali-metal atoms. *Phys. Rev. A* **39**, 5613–5623 (1989).
- Xu, S., Sha, G. & Xie, J. Cavity ring-down spectroscopy in the liquid phase. *Rev. Sci. Instrum.* **73**, 255–258 (2002).
- Fini, J. M. Microstructure fibres for optical sensing in gases and liquids. *Meas. Sci. Technol.* **15**, 1120–1128 (2004).
- Cox, F. M., Argyros, A. & Large, M. C. J. Liquid-filled hollow core microstructured polymer optical fiber. *Opt. Expr.* **14**, 4135–4140 (2006).
- Maystre, F. & Bertholds, A. Magneto-optic current sensor using a helical-fiber Fabry-Perot resonator. *Opt. Lett.* **14**, 587–589 (1989).
- Olson, D. L., Peck, T. L., Webb, A. G., Magin, R. L. & Sweedler, J. V. High-resolution microcoil ^1H -NMR for mass-limited, nanoliter-volume samples. *Science* **270**, 1967–1970 (1995).

Acknowledgements This work was supported by the US DOE and the US NSF. We thank W. Happer's group for use of the 9 T superconducting magnet and S. Smullin for comments on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.V.R. (romalis@princeton.edu).

Controls on tropical Pacific Ocean productivity revealed through nutrient stress diagnostics

Michael J. Behrenfeld¹, Kirby Worthington², Robert M. Sherrell³, Francisco P. Chavez⁴, Peter Strutton⁵, Michael McPhaden⁶ & Donald M. Shea⁷

In situ enrichment experiments have shown that the growth of bloom-forming diatoms in the major high-nitrate low-chlorophyll (HNLC) regions of the world's oceans is limited by the availability of iron^{1–3}. Yet even the largest of these manipulative experiments represents only a small fraction of an ocean basin, and the responses observed are strongly influenced by the proliferation of rare species rather than the growth of naturally dominant populations^{4,5}. Here we link unique fluorescence attributes of phytoplankton to specific physiological responses to nutrient stress, and use these relationships to evaluate the factors that constrain phytoplankton growth in the tropical Pacific Ocean on an unprecedented spatial scale. On the basis of fluorescence measurements taken over 12 years, we delineate three major ecophysiological regimes in this region. We find that iron has a key function in regulating phytoplankton growth in both HNLC and oligotrophic waters near the Equator and further south, whereas nitrogen and zooplankton grazing are the primary factors that regulate biomass production in the north. Application of our findings to the interpretation of satellite chlorophyll fields shows that productivity in the tropical Pacific basin may be 1.2–2.5 Pg C yr⁻¹ lower than previous estimates have suggested, a difference that is comparable to the global change in ocean production that accompanied the largest El Niño to La Niña transition on record⁶.

The tropical Pacific is characterized by warm, well-stratified and nutrient-poor waters separated by a major upwelling plume of nutrient-rich water near the Equator extending from roughly the dateline (180°) to the eastern boundary (Fig. 1a). The upwelled water is rich in dissolved CO₂ that subsequently degasses to the atmosphere. Although the region is the largest natural oceanic source of CO₂ to the atmosphere^{7,8}, the extent of this CO₂ release is curtailed to about 0.7–1.5 Pg yr⁻¹ (1 Pg = 10¹⁵ g) by the photosynthetic carbon uptake of an elevated phytoplankton biomass supported by upwelled macronutrients and micronutrients^{8,9}. Throughout the tropical Pacific, variations in physical and chemical properties of the upper ocean imprint resident phytoplankton with physiological characteristics diagnostic of their specific growth constraints. These physiological expressions can be distinguished by associated diel patterns in normalized variable fluorescence (F_v/F_m)¹⁰.

We collected more than 140,000 measurements of variable fluorescence along 58,000 km of ship transects during ten field studies between 1994 and 2006 to characterize the broad-scale biological and physiological features of the tropical Pacific (Fig. 1a). Phytoplankton biomass in the study area is distributed similarly to surface nitrate (Fig. 1a), with low values north of roughly 9°N and west of the

dateline, and distinctly elevated values in the equatorial upwelling zone. Daily F_v/F_m patterns in the tropical Pacific exhibit three dominant features: first, maxima at sunrise and sunset; second, a midday suppression from photoinhibition; and third, a nocturnal decrease (Fig. 1b, c). We find dawn maxima in F_v/F_m to vary inversely with biomass, having high values in low-nitrate areas and decidedly lower values in upwelling waters (Figs 1b and 2a). Midday photoinhibition

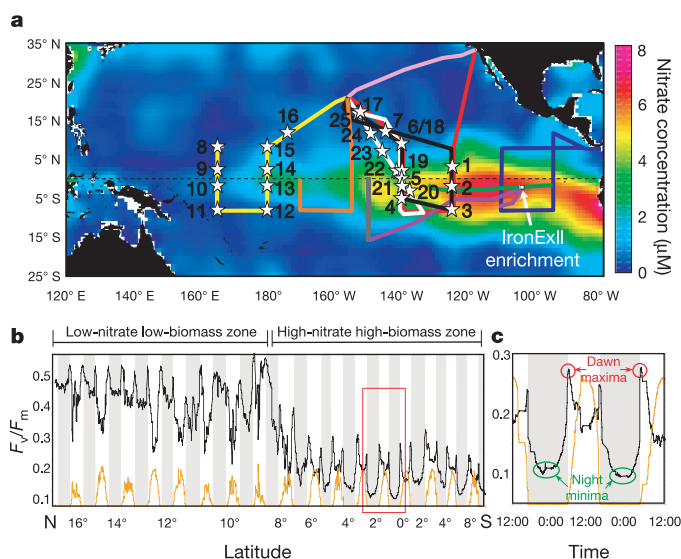


Figure 1 | The tropical Pacific study area. **a**, Ship transects (lines) and enrichment experiment locations (stars) for the ten field studies. Transect dates are as follows: grey, 1994; dark purple, 1995; black, 2000; yellow, pink, orange and green, 2001; white, 2002; blue, 2003; red, 2006. The background colour is Levitus climatological surface nitrate concentration, showing elevated values in the equatorial upwelling zone. **b**, Typical variations in normalized variable fluorescence (F_v/F_m) for the tropical Pacific, with high dawn values and large photoinhibition in low-nitrate low-biomass regions and low dawn values and large nocturnal decreases in high-nitrate high-biomass areas (data from the 2000 transect, corresponding to the black line in **a** extending from Hawaii to 8°S, 140°W). **c**, Primary diel features in F_v/F_m (from the red box in **b**) are a nocturnal decrease (circled in green), a dawn maximum (circled in red) and midday photoinhibition. To illustrate the scale of the current investigation, the 64 km² IronExII experiment of 1996, which dwarfed earlier bottle enrichments, was only 1/100 the size of the small white rectangle shown in **a** at 2°S, 104°W. The orange lines in **b** and **c** indicate incident sunlight (relative).

¹Department of Botany and Plant Pathology, Cordley Hall 2082, Oregon State University, Corvallis, Oregon 97331-2902, USA. ²National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. ³Institute of Marine and Coastal Sciences and Department of Geological Sciences, Rutgers University, 71 Dudley Road, New Brunswick, New Jersey 08901-8521, USA. ⁴Monterey Bay Aquarium Research Institute, 7700 Sandholdt Road, Moss Landing, California 95039-9644, USA. ⁵College of Oceanic and Atmospheric Sciences, 104 COAS Admin Building, Oregon State University, Corvallis, Oregon 97331-5503, USA. ⁶National Oceanic and Atmospheric Administration Pacific Marine Environmental Laboratory, 7600 Sand Point Way NE, Seattle, Washington 98115, USA. ⁷Science Applications International Corporation, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA.

is also enhanced in regions of low biomass (Fig. 1b). The remarkable nocturnal decreases in F_v/F_m are uniform in the upper water column and then disappear in the light-limited lower reaches of the photic zone (see Supplementary Information).

We delineated three physiological regimes in the tropical Pacific on the basis of dawn maxima in F_v/F_m and the extent of the nocturnal F_v/F_m decrease (Fig. 2). Regime I is largely found in the north and is characterized by high values (0.45 or more) of F_v/F_m at dawn and small (less than 25%) nocturnal decreases. Regime II is found west of the upwelling plume and has similarly elevated dawn maxima (Fig. 2a), but exhibits large nocturnal decreases in F_v/F_m (Fig. 2b). Regime III is distinguished by having chronically low dawn F_v/F_m and pronounced decreases in F_v/F_m at night (Fig. 2). The link between these three physiological regimes and specific nutritional constraints was directly tested by conducting 25 small-volume (10-litre) nutrient enrichment experiments (Fig. 1a). These 21–36-h experiments included representatives from all three regimes and involved enrichments of 5 μM NO_3 , 5 μM NH_4 , 1 μM PO_4 and 4 nM iron (see Supplementary Information). Their short duration ensured that observed responses reflected immediate physiological changes rather than growth of the dominant species or a bloom of a rare species.

The addition of NO_3 , NH_4 or PO_4 had no significant influence on photosynthetic characteristics of samples from regime III. In contrast, addition of iron to samples from regime III consistently eliminated the nocturnal decrease in F_v/F_m , markedly increased overall F_v/F_m values to regimes I and II levels (Fig. 3a), enhanced functional absorption cross-sections of oxygen-evolving photosystem II (PSII) complexes and caused a major decrease in electron

turnover times of the plastoquinone (PQ) pool (Fig. 3b, c). In regime II, iron addition similarly removed the nocturnal decrease in F_v/F_m and generally enhanced electron turnover rates, but dawn F_v/F_m values were already near maximal and did not change with iron (or PO_4) enrichment (Fig. 3a). Additions of NO_3 and NH_4 caused the most striking responses in regime II, inducing marked decreases in F_v/F_m (Fig. 3a) that were associated with increased background fluorescence levels. Finally, photosynthetic characteristics in regime I remained unaltered after the addition of iron or PO_4 , and none of the nutrient treatments significantly influenced PSII absorption cross-sections or PQ electron turnover (Fig. 3). The only clear physiological response in regime I was a decrease in F_v/F_m during four of the eight experiments after NO_3 or NH_4 amendment (Fig. 3a). Again, these decreases in F_v/F_m resulted from an increase in background fluorescence.

Clearly, the three tropical Pacific regimes correspond to different conditions of iron and nitrogen availability. Because many details on the effects of iron stress in plants are known, our population-level fluorescence diagnostics can now be associated with specific phenomena in photosynthetic membranes. Importantly, the photosynthetic apparatus is a major sink for cellular iron, with 24 iron

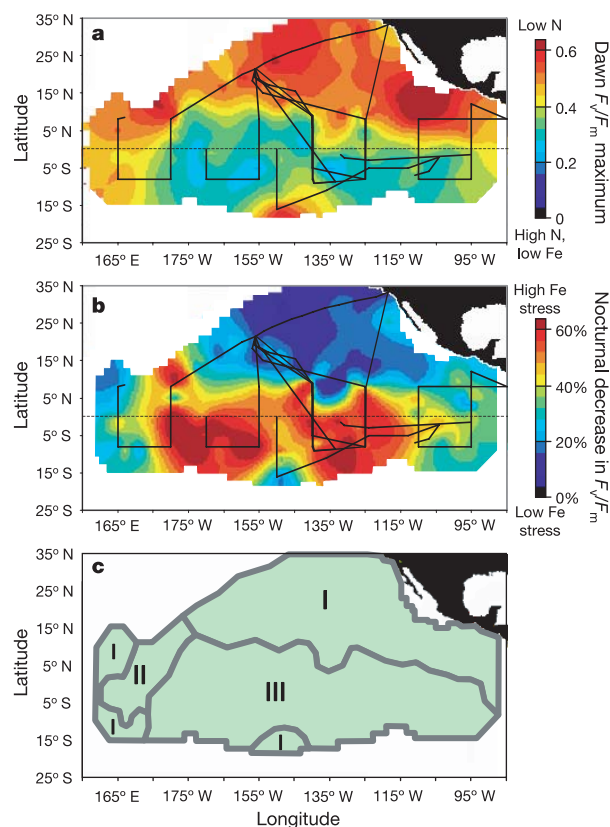


Figure 2 | Fluorescence diagnostics delineate three physiological regimes in the tropical Pacific. **a**, Dawn F_v/F_m maximum. **b**, Nocturnal decrease in F_v/F_m . **c**, The three regimes: regime I (small nocturnal decreases and high dawn values of F_v/F_m), regime II (large nocturnal decreases and high dawn values of F_v/F_m) and regime III (large nocturnal decreases and low dawn values of F_v/F_m).

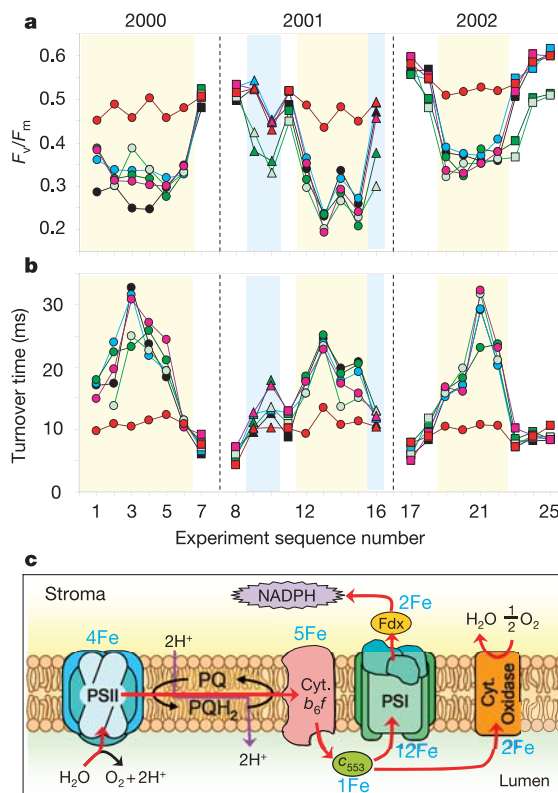


Figure 3 | Nutrient enrichment responses in the three physiological regimes. **a**, **b**, Initial and end-of-experiment treatment values of normalized variable fluorescence (F_v/F_m) (**a**) and plastoquinone pool electron turnover times during the 25 enrichment experiments (stars in Fig. 1a) (**b**). Regime I is represented by squares and no background shading; regime II by triangles and blue shading, and regime III by circles and yellow shading. Black symbols and lines, zero time control; blue, end-of-experiment control; dark green, NO_3 ; light green, NH_4 ; pink, PO_4 ; red, iron. Vertical dashed lines separate cruises (labelled at the top). Experiment sequence number (x axis) follows Fig. 1a. **c**, Primary components of a photosynthetic membrane and their iron requirements (blue text). The photosynthetic electron transport sequence is as follows: PSII to plastoquinone pool (PQ, PQH_2) to cytochrome b_6f to a mobile cytochrome (c_{553}) and finally to either PSI and ferredoxin (Fdx) or a terminal cytochrome oxidase. In prokaryotes and eukaryotes, electron transport can also proceed directly from the PQ pool to terminal oxidases.

atoms required for a single complete copy of the electron transport chain (Fig. 3c)¹¹. Iron stress significantly depletes photosynthetic components on the acceptor side of the PQ pool (cytochrome *b₆f*, cytochrome oxidase, photosystem I (PSI) and ferredoxin) (Fig. 3c)^{12–15}. These changes are responsible for the large nocturnal F_v/F_m decreases in regimes II and III.

All phytoplankton use their photosynthetic membranes for respiratory electron transport in the dark (termed 'chlororespiration' in eukaryotes) (Fig. 3c). Consequently, the PQ pool is generally mildly reduced at night, but under iron-stressed conditions this reduction is severe because electron transport is rate-limited by diminished cytochrome concentrations (Fig. 3c)^{10,12,15–18}. Because the redox state of the primary electron acceptor on PSII (Q_A) exists in equilibrium with that of the PQ pool¹⁷, severe reduction of PQ at night results in a back-transfer of electrons to Q_A that we detected as a decrease in F_v/F_m . Being a respiration-driven phenomenon, the degree of PQ pool reduction depends on the pool size of respiratory substrates. Depletion of these substrates is responsible for the modest increase in F_v/F_m each night before sunrise (Fig. 1c) and explains the observation that the nocturnal decrease in F_v/F_m cannot be artificially induced after sunrise by re-exposure to darkness¹⁷ until an adequate photosynthate pool has been rebuilt. The immediate increase in F_v/F_m at dawn results from oxidation of the PQ pool by PSI turnover (Fig. 1c), a response that can be artificially replicated at night by exposure to PSI-specific light^{10,17}. Rapid loss of the nocturnal decrease in F_v/F_m and increased electron turnover (Fig. 3b) in regimes II and III on the addition of iron reflect an associated induction of cytochrome synthesis^{12,16}.

The distinguishing characteristic of regime III is low dawn F_v/F_m maxima (Fig. 2a). This feature is caused by unique chlorophyll–protein complexes that are synthesized when reduced nitrogen is abundant but iron is limiting^{12,18–23}. These complexes seem to function in a photoprotective manner, are to a large degree functionally 'disconnected' from PSII, and have high background fluorescence that decreases F_v/F_m (refs 11, 12, 18, 22–24). Pigment from the

complexes is immediately transferred to functional PSII antennae on iron enrichment, leading to decreased background fluorescence and increases in F_v/F_m (Fig. 3a) and PSII absorption cross-sections. This same response is regularly observed in laboratory iron-recovery experiments^{12,16,20,25,26}, proceeds in the presence of chlorophyll synthesis inhibitors²⁶ but not protein synthesis inhibitors¹⁵, and suggests that the iron-induced complexes function as pigment reservoirs to facilitate recovery from iron stress^{12,20,25}. In regime II, the process is simply reversed by the addition of NO_3 or NH_4 , which induces the synthesis of these special pigment–protein complexes that are naturally substrate limited by nitrogen availability. The occasional appearance of this 'regime II-type' response in regime I experiments (Fig. 3a) indicates that these northern waters, which are regulated by nitrogen and grazing, can easily be perturbed into iron-stressed conditions²⁷.

Our two variable fluorescence diagnostics define four possible physiological regimes, three of which are present in the tropical Pacific (Fig. 4). High dawn F_v/F_m values occur in both nitrogen-limited and iron-limited systems, as long as reduced nitrogen levels are low (regimes I and II). Low dawn F_v/F_m values result when iron is limiting and elevated nitrogen levels allow the synthesis of the special pigment–protein complexes (classic HNLC conditions). Large nocturnal decreases in F_v/F_m require both iron stress and sufficient dark respiration to drive PQ pool reduction (regimes II and III). Conversely, small nocturnal decreases in F_v/F_m result when iron stress is absent or growth is too slow to cause PQ pool reduction at night. Accordingly, the fourth physiological regime, which was not found in the tropical Pacific and is defined by low dawn F_v/F_m and small nocturnal decreases (Fig. 4), is predictably observed in polar HNLC regions where iron is limiting and nitrate is replete, but growth rates are too low for significant night-time PQ pool reduction²⁸.

The tropical Pacific basin is responsible for roughly 20% (9–14 Pg C yr⁻¹) of global ocean productivity (see Supplementary Information) and has a prominent function in air–sea CO₂ exchange. Assessing constraints on productivity in this permanently stratified region is challenging because the dominant phytoplankton are naturally growing at relatively high rates (of the order of one division per day) and their standing stock changes little with nutrient enrichment⁴. Physiological diagnostics provide a solution to this problem and here are applied on an unprecedented scale. Special pigment–protein complexes synthesized during iron stress underlie one of our key fluorescence attributes. These structures cause an enhanced greenness in HNLC regions (that is, regime III) that is not associated with elevated photosynthesis. This effect is quantitatively related to the suppression of dawn F_v/F_m values and must be accounted for when satellite surface chlorophyll fields are used to estimate ocean productivity or to evaluate ocean-circulation–ecosystem model performance. We assessed the influence of these iron-induced structures by adjusting satellite chlorophyll fields with our variable fluorescence data from the field; we found tropical Pacific production to be 1.2–2.5 Pg C yr⁻¹ lower than for uncorrected fields (see Supplementary Information). This difference is comparable to global productivity changes during major El Niño to La Niña transitions (0.4–0.8 Pg C yr⁻¹)^{7,29} and underscores the importance of characterizing nutrient constraints and their physiological consequences. Our current climatological treatment unquestionably misses seasonal and interannual variations in boundaries between physiological regimes. Future evaluations will benefit if links can be established between nutritional conditions and remote sensing properties, such as Moderate Resolution Imaging Spectrometer (MODIS) solar-stimulated fluorescence.

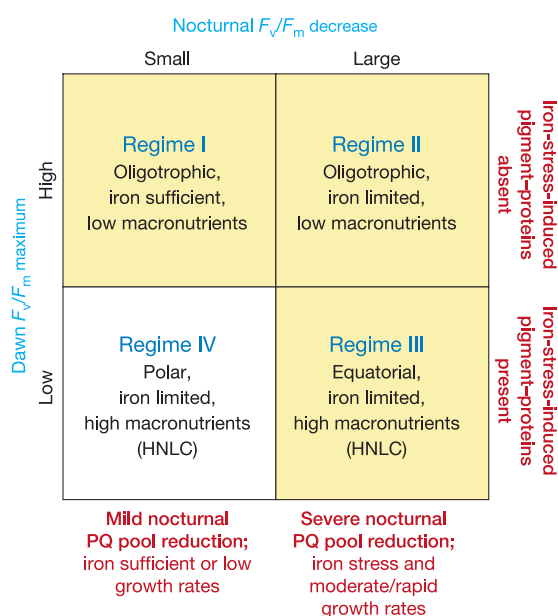


Figure 4 | Environmental conditions corresponding to the four physiological regimes. The regimes are distinguished by dawn normalized variable fluorescence values (F_v/F_m) (low, less than 0.45; high, 0.45 or more) and the percentage decrease in F_v/F_m at night (small, 25% or less; large, more than 25%). Physiological and nutritional conditions responsible for the diagnostic attributes are summarized in red text to the right and at the bottom of the matrix. The three regimes observed in the tropical Pacific are identified by a yellow background.

METHODS

The Supplementary Information contains details of all experimental methods and model calculations.

Fluorescence. A fast-repetition-rate fluorimeter (FRRF) was used to measure chlorophyll fluorescence characteristics of phytoplankton continuously sampled

from the surface mixed layer. The FRRf protocol involves a rapid sequence of subsaturating light flashes that cause a rise in fluorescence *in vivo* from an initial (F_0) to a maximal (F_m) level. This change in fluorescence (F_v) is associated with absorbed light energy used for photosynthesis and is normalized to F_m (that is, F_v/F_m) to account for biomass variability. Functional absorption cross-sections of PSII can also be derived from the rate of increase between F_0 and F_m . Electron transport rates downstream of PSII are determined from fluorescence decay kinetics after a saturating sequence of light flashes.

Received 14 December 2005; accepted 19 July 2006.

- Coale, K. *et al.* A massive phytoplankton bloom induced by an ecosystem-scale iron fertilization experiment in the equatorial Pacific ocean. *Nature* **383**, 495–501 (1996).
- Boyd, P. W. *et al.* A mesoscale phytoplankton bloom in the polar Southern ocean stimulated by iron fertilization. *Nature* **407**, 695–702 (2000).
- Tsuda, A. *et al.* A mesoscale iron enrichment in the western Subarctic Pacific induces a large centric diatom bloom. *Science* **300**, 958–961 (2003).
- Cavender-Bares, K. K., Mann, E. L., Chisholm, S. W., Ondrusek, M. E. & Bidigare, R. R. Differential response of equatorial Pacific phytoplankton to iron fertilization. *Limnol. Oceanogr.* **44**, 237–246 (1999).
- Olson, R. J., Sosik, H. M., Chekalyuk, A. M. & Shalapyonok, A. Effects of iron enrichment on phytoplankton in the Southern ocean during late summer: active fluorescence and flow cytometric analyses. *Deep-sea Res. II* **47**, 3181–3200 (2000).
- Behrenfeld, M. J. *et al.* Biospheric primary production during an ENSO transition. *Science* **291**, 2594–2597 (2001).
- Takahashi, T. *et al.* Global sea-air CO₂ flux based on climatological surface ocean pCO₂, and seasonal biological and temperature effects. *Deep-sea Res. II* **49**, 1601–1622 (2002).
- Chavez, F. P. *et al.* Biological and chemical response of the equatorial Pacific ocean to the 1997–98 El Niño. *Science* **286**, 2126–2131 (1999).
- Feely, R. A., Wanninkhof, R., Takahashi, T. & Tans, P. Influence of El Niño on the equatorial Pacific contribution to atmospheric CO₂ accumulation. *Nature* **398**, 597–601 (1999).
- Behrenfeld, M. J. & Kolber, Z. S. Widespread iron limitation of phytoplankton in the south Pacific ocean. *Science* **283**, 840–843 (1999).
- Michel, K.-P. & Pistorius, E. K. Adaptation of the photosynthetic electron transport chain in cyanobacteria to iron deficiency: The function of *IdiA* and *IsiA*. *Physiol. Plant.* **120**, 36–50 (2004).
- Vassiliev, I. R. *et al.* Effects of iron limitation on photosystem II composition and light utilization in *Dunaliella tertiolecta*. *Plant Physiol.* **109**, 963–972 (1995).
- La Roche, J., Boyd, P. W., McKay, R. M. L. & Geider, R. J. Flavodoxin as an *in situ* marker for iron stress in phytoplankton. *Nature* **382**, 802–805 (1996).
- Ivanov, A. G. *et al.* Iron stress restricts photosynthetic intersystem electron transport in *Synechococcus* sp. PCC 7942. *FEBS Lett.* **485**, 173–177 (2000).
- Sandmann, G. Consequences of iron deficiency on photosynthetic and respiratory electron transport in blue-green algae. *Photosynth. Res.* **6**, 261–271 (1985).
- Greene, R. M., Geider, R. J., Kolber, Z. & Falkowski, P. G. Iron-induced changes in light harvesting and photochemical energy conversion processes in eukaryotic marine algae. *Plant Physiol.* **100**, 565–575 (1992).
- Belkhdja, R. *et al.* Iron deficiency causes changes in chlorophyll fluorescence due to the reduction in the dark of the photosystem II acceptor side. *Photosynth. Res.* **56**, 265–276 (1998).
- Morales, F., Moise, N., Quílez, R., Abadía, A. & Moya, I. Iron deficiency interrupts energy transfer from a disconnected part of the antenna to the rest of photosystem II. *Photosynth. Res.* **70**, 207–220 (2001).
- Guikema, J. A. & Sherman, L. A. Organization and function of chlorophyll in membranes of cyanobacteria during iron starvation. *Plant Physiol.* **73**, 250–256 (1983).
- Reithman, H. C. & Sherman, L. A. Purification and characterization of an iron stress-induced chlorophyll-protein from the cyanobacterium *Anacystis nidulans* R2. *Biochim. Biophys. Acta* **935**, 141–151 (1988).
- Varsano, T., Kaftan, D. & Pick, U. Effects of iron deficiency on thylakoid membrane structure and composition in the alga *Dunaliella salina*. *J. Plant Nutr.* **26**, 2197–2210 (2003).
- Moseley, J. L. *et al.* Adaptation to Fe-deficiency requires remodeling of the photosynthetic apparatus. *EMBO J.* **21**, 6709–6720 (2002).
- Sandström, S., Ivanov, A. G., Park, Y.-I. I., Öquist, G. & Gustafsson, P. Iron stress responses in the cyanobacterium *Synechococcus* sp. PCC7942. *Physiol. Plant.* **116**, 255–263 (2002).
- Park, Y.-I. I., Sandström, S., Gustafsson, P. & Öquist, G. Expression of the *isiA* gene is essential for the survival of the cyanobacterium *Synechococcus* sp. PCC 7942 by protecting photosystem II from excess light under iron limitation. *Mol. Microbiol.* **32**, 123–129 (1999).
- Troyan, T. A., Bullerjahn, G. S. & Sherman, L. A. in *Techniques and New Developments in Photosynthesis Research* (eds Barber, J. & Malkin, R.) 601–604 (Plenum Press, New York, 1989).
- Larbi, A., Abadía, A., Morales, F. & Abadía, J. Fe resupply to Fe-deficient sugar beet plants leads to rapid changes in the violaxanthin cycle and other photosynthetic characteristics without significant *de novo* chlorophyll synthesis. *Photosynth. Res.* **79**, 59–69 (2004).
- DiTullio, G. R., Hutchins, D. A. & Bruland, K. W. Interaction of iron and major nutrients controls phytoplankton growth and species composition in the tropical North Pacific Ocean. *Limnol. Oceanogr.* **38**, 495–508 (1993).
- Boyd, P. W. & Abraham, E. R. Iron-mediated changes in phytoplankton photosynthetic competence during SOIREE. *Deep-sea Res. II* **48**, 2529–2550 (2001).
- Behrenfeld, M. J., Boss, E., Siegel, D. A. & Shea, D. M. Carbon-based ocean productivity and phytoplankton physiology from space. *Glob. Biogeochem. Cycles* **19**, GB1006, doi:10.1029/2004GB002299 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Z. Kolber, R. O'Malley, and the crew and officers of the NOAA ships *Ka'imimoana* and *Ronald Brown*. This research was funded by the National Science Foundation, the National Aeronautics and Space Administration, and the National Oceanic and Atmospheric Administration's Tropical Atmosphere Ocean array programme.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.J.B. (mjb@science.oregonstate.edu).

Figure 2 shows images from four seismic reflection lines over the centre of the volcano. The southern line 8, over the flattest part of the volcano, shows a 4-km-wide reflection event at ~ 1.3 s below the sea floor. The polarity of this reflection is the reverse of that of the sea floor and layer 2A reflections (Fig. 3a), suggesting that it is produced by a negative velocity contrast. We suggest that this is an axial magma chamber (AMC) reflection. If we assume an average velocity of 5 km s^{-1} in the upper crust, then this reflector lies 3.25 km below the sea floor. We also observe a layer 2A reflection at about 0.5 s below the sea floor (1 km depth for an assumed velocity of 4 km s^{-1}). We notice reflection events that dip towards the centre of the ridge and are roughly aligned with the axial valley bounding fault-line scarps observed on the sea floor, which we interpret as rift bounding faults (East Bounding Fault (EBF) and West Bounding Fault (WBF), Fig. 2). Smaller faults that bound the graben within the rift valley floor are also visible in the proximity of the Lucky Strike hydrothermal field, and reach the base of layer 2A. To visualize the geometry of the faults in the crust, we remove the seafloor effect by flattening the seismic image using a smooth seafloor topography (Supplementary Fig. S1), which clearly shows that these faults arrive at the sea floor. The event on the eastern side (EBF) is strong and can be observed on common mid-point data (gather) (Fig. 3c). If we assume an average velocity of 5 km s^{-1} , the fault dip is $45\text{--}50^\circ$, consistent with dips determined from earthquake fault plane solutions at slow-spreading ridges¹⁸. The event seems to continue 0.6 s below the AMC, but this could be an imaging effect that may disappear after migration. However, we can confidently say that the EBF goes down to at least the AMC depth. The event on the western side (WBF) has a shallower dip and does not appear to continue below the AMC (Fig. 2). We also observe a steep event (50°) on the western flank of the volcano that terminates above the AMC. This event is associated with the western fault bounding the graben on top of the volcano.

Line 20 crosses the centre of the volcano, where the hydrothermal vents are observed. Layer 2A is slightly thicker (0.55 s) and the AMC is slightly narrower (3.5 km) than those along line 8. The dipping events interpreted as faults are also present on this line. The hydrothermal field lies around small inward-dipping faults that go down to the bottom of layer 2A. Line 37 traverses the northern, rifted part of the volcano. The layer 2A is thinnest (0.5 s) beneath the graben floor and thickest (0.65 s) beneath the volcanic ridge. The AMC appears to narrow northward from 4 km to 2.5 km over a distance of 2.8 km. A well-imaged westward-dipping event, interpreted as a fault, penetrates downward from the eastern graben wall.

Line 1011 is an along-axis profile that extends over the entire volcanic edifice (Figs 1 and 2). The layer-2A reflection lies about 0.5 s below the sea floor, except beneath the volcano where it is slightly thicker (0.65 s). The AMC is observed for about 7 km along the axis. We also observe several subhorizontal features below layer 2A, which could be caused by frozen melt sills or ridge-parallel faults, such as the graben bounding faults.

The presence of an AMC has been reported along the Reykjanes ridge using seismic reflection and refraction data⁸. Although the Reykjanes ridge is a part of the MAR, it lacks slow-spreading-ridge characteristics and has been influenced by the nearby Iceland hot-spot. A small AMC reflection was also reported beneath the Snake Pit hydrothermal field at the MAR, but the quality of the image is poor¹⁹. Seismic tomography results have shown the presence of low velocities in the lower crust indicative of high temperatures and partial melt²⁰, but no AMC reflections were observed.

The observation of a large AMC at crustal depths beneath the Lucky Strike segment has important consequences for our understanding of magmatic-tectonic interactions at slow-spreading ridges. Figure 4 shows a schematic diagram summarizing the three-dimensional

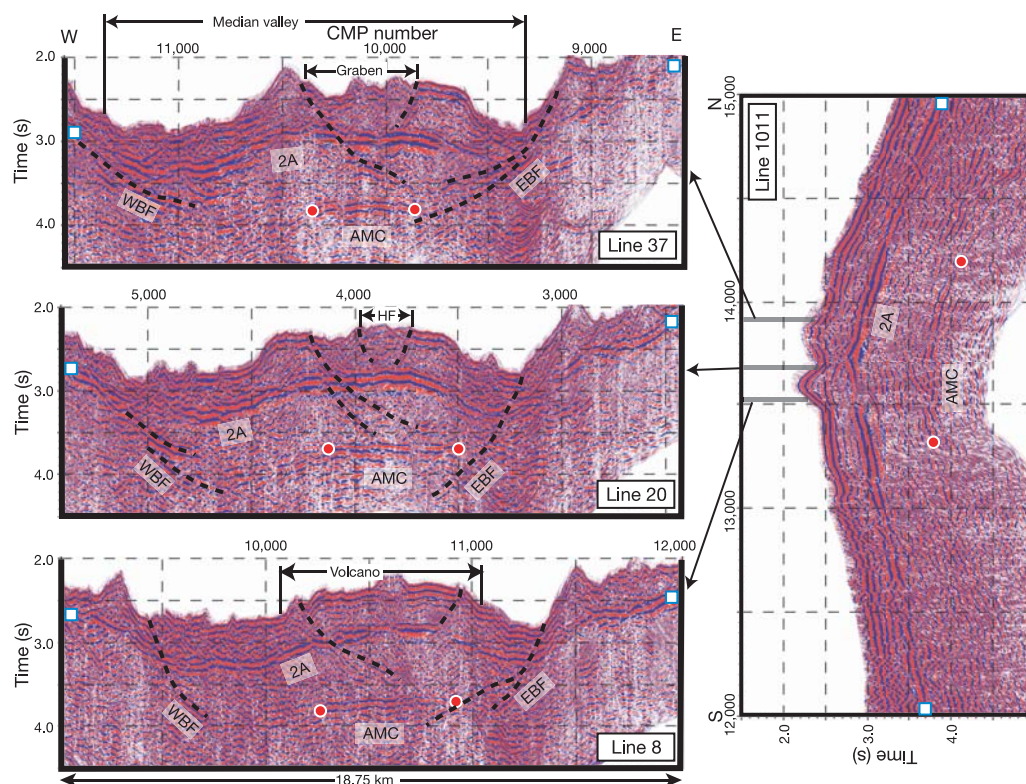


Figure 2 | Seismic reflection images of the AMC and faults. Seismic sections across the ridge axis (lines 8, 20 and 37) and along the ridge axis (line 1011). The vertical exaggeration is 1:1, assuming a velocity of 5 km s^{-1} . Blue and red ripples represent positive and negative trace amplitudes, respectively. Red circles show the limits of the AMC reflector, blue squares mark the layer 2A reflector and dashed black lines indicate principal faults

(Supplementary Fig. S2 is without interpretive curves). Labels are placed just beneath the reflector to which they correspond. The AMC width decreases from the south (4 km) to the north (2.5 km) but the AMC continues for about 3 km north of the three-dimensional box (line 1011), beneath the graben. HF, hydrothermal vent field. CMP, common mid-point.

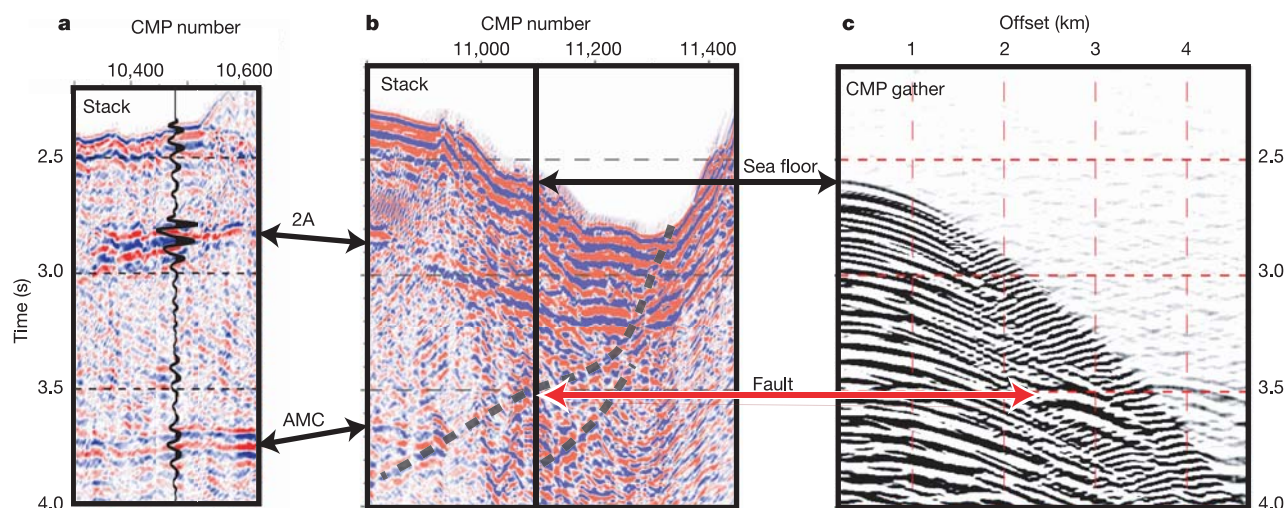


Figure 3 | Fault and AMC reflector details. Detailed seismic sections from line 8. **a**, Section showing the sea floor, layer 2A and the AMC reflector. The AMC reflector has the opposite polarity of the other reflectors. The black wiggle trace highlights the polarity of different events. **b**, The EBF, enlarged

from Fig. 1. **c**, Super-CMP gather corresponding to the vertical black line in **b**, showing the fault reflection. This gather was obtained by combining six CMP and has been frequency-wavenumber (F-K) filtered to highlight the fault reflection.

geometry of the magma chamber, faults, and the seafloor features. Axial seismicity²¹ and thermo-mechanical models of axial valley formation²² suggest that isotherms corresponding to the brittle–ductile transition (500–700 °C) may lie as deep as 8 km below the sea floor at slow-spreading ridges. The presence of rift-bounding faults that seem to go down to at least 3–4 km depth suggests that the rift-forming tectonic processes (faulting) have coexisted with magmatism in the recent past. It is difficult to establish the age of the different structures, but it is certain that the graben bounding faults are younger than the volcano, because they cross-cut its otherwise unfaulted and smooth flanks. The synchronous presence of the AMC and deep penetrating faults suggests that there is a continuous interplay between magmatic and tectonic processes.

The discovery of the crustal magma chamber and deep penetrating faults also has important implications for the crustal accretion process at slow-spreading ridges. Thick crust beneath the volcano^{9,14} may result from a robust melt supply at the centre of the segment where the crust is formed by cooling and crystallization of magma in crustal melt lenses²³, while lateral propagation of dykes over tens of kilometres could be responsible for crustal accretion towards the segment ends²⁴. The neighbouring Azores hotspot could be responsible for higher melt production in the mantle, which could get focused at the centre of the segment, sustaining a long-lived magma

chamber in the crust. On the other hand, the overall morphology of the Lucky Strike segment with a well-developed axial valley suggests that the crust is largely cooled and brittle, and supports the idea that the imaged AMC might be the first direct evidence of melt at the roof of a short-lived crustal melt body. The composition of gabbros and overall lithospheric structure from slow-spreading ridges²⁵ does suggest that gabbros crystallize in intrusive bodies within a tectonically deforming crust and shallow mantle²⁶. However, the high regional bathymetry, the presence of a large central volcano and a crustal AMC lead us to suggest that magmatism dominates the crustal accretion process at the Lucky Strike segment of the MAR.

It has been suggested that high-temperature seafloor hydrothermal activities at fast-spreading centres are linked to fresh melt supply in the crust²⁷. The high-temperature hydrothermal vents observed on the Lucky Strike volcano also require a magmatic heat source, and we suggest that the AMC observed on the seismic profiles is this source. The axial valley bounding faults appear to penetrate down to AMC depth, so they may act as recharge pathways for hydrothermal circulation, allowing for sea water to penetrate efficiently down to near the AMC. Heated fluids would then flow vertically upwards in a narrow zone, forming the hydrothermal vent field²⁸.

Our results provide the first image, to our knowledge, of an axial magma chamber at a slow-spreading ridge segment, and its

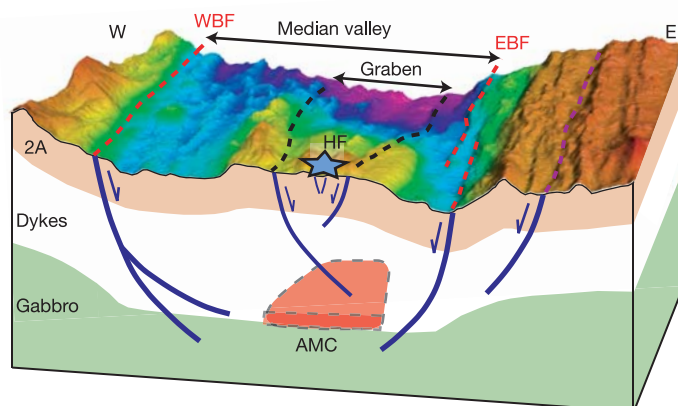


Figure 4 | Three-dimensional schematic view of the AMC and faults.

Schematic diagram relating different features observed on seismic profiles to the bathymetry. Beneath the Lucky Strike volcano, layer 2A is about 1 km thick and a large AMC is present (up to 7 km long and 4 km wide). The median valley faults seem to continue down to and possibly beneath the AMC. The graben bounding faults terminate above the AMC in dykes and layer 2A.

interaction with active tectonic and hydrothermal processes. The along-axis extension of the magma chamber (<10 km) compared to the total segment length (~ 60 km) suggests a very efficient lateral dyke propagation away from the AMC, a length similar to that of dyking events at intermediate-spreading ridges²⁹. Our results are also consistent with the presence of melt in the lower crust²⁰, and possible seismic reflectors that may be associated with melt bodies under the Snake Pit vent field¹⁹. We expect magma chambers to be common under long-lived hydrothermal fields at slow-spreading segments, particularly those within the rift valley, at or near the centre of segments, and hosted in basalt (such as Menez Gwen, Snake Pit³⁰).

METHODS

Seismic reflection data were acquired using a 4.5-km-long digital streamer with 12.5-m receiver intervals towed at 15 m depth. The energy source was a 2,594 cubic inch 18-airgun array towed at a depth of about 12 m and tuned to provide a broadband (8–50 Hz) source with a strong first arrival. The record length was 11 s and the sample interval was 2 ms. A set of thirty-nine 18.75-km-long seismic reflection lines were spaced at 100 m apart and provided three-dimensional coverage in an 18.75×3.8 km² area. The shot interval was 37.5 m and the vessel speed was 4.5 knots. To obtain a high-resolution lateral image (6.25 m) across the ridge axis, lines were shot perpendicular to the ridge (109° azimuth). The along-axis line 1011 was shot at an interval of 75 m using a 5,638 cubic inch airgun array.

The seismic reflection data were quality controlled using nearly real-time onboard processing. The four lines (8, 20, 37, 1011) presented here were reprocessed assuming a two-dimensional geometry. The data were re-sampled to 4 ms and binned at 6.25 m spacing. A set of constant-velocity stacks was calculated to obtain stacking velocities for layer 2A and the AMC. The image quality is very sensitive to velocity so several tests were performed to optimize the velocity. Layer 2A stacked best at $1,900\text{--}2,100$ m s⁻¹, whereas the AMC stacked at $3,000\text{--}3,300$ m s⁻¹. A dipping reflector such as a fault requires higher stacking velocities. The median valley faults stacked best at a velocity of $3,000\text{--}4,000$ m s⁻¹, depending upon the dip and depth. To relate images from one line to another and remove bias due to velocity model errors, we decided to use the same crustal velocities for all the lines: $1,900$ m s⁻¹ down to 600 ms beneath the sea floor and $3,100$ m s⁻¹ below that level. Slightly higher velocity does not influence the imaging of the sea floor because only near-offset data contribute to the stack as stretched large-offset data get muted (deleted) after normal moveout correction.

Seafloor scattering and multiples were the strongest noise in the data. To reduce the effect of multiples during migration, the seafloor multiples were muted. To remove the seafloor scattering, we used a frequency domain migration with a constant velocity of $1,500$ m s⁻¹. Migration is sensitive to velocity, and therefore a constant velocity migration provided uniform and objective images. When interpreting these images, one should note that stacking flattens a dipping interface, and therefore after full crustal migration, a dipping event will get steeper and will map to a shallower depth.

There were several factors that facilitated the imaging of the magma chamber and faults. The fine lateral sampling (6.25 m), high fold (60) and the broadband source were essential. The line spacing of 100 m provided the along-axis continuity of the events. The noise due to seafloor scattering was reduced because of the smooth bathymetry over the Lucky Strike volcano and the presence of a wide median valley (12–15 km). The processing sequence used here was extremely valuable for preserving the lateral continuity of events along the line and correlation of events on different lines.

Received 17 March; accepted 18 July 2006.

1. Cannat, M. Emplacement of mantle rocks in the seafloor at mid-ocean ridges. *J. Geophys. Res.* **98**, 4163–4172 (1993).
2. Collier, J. S. & Sinha, M. C. Seismic images of a magma chamber beneath the Lau Basin back-arc spreading centre. *Nature* **346**, 646–648 (1990).
3. Canales, J. P. *et al.* Upper crustal structure and axial topography at intermediate spreading ridges: Seismic constraints from the southern Juan de Fuca Ridge. *J. Geophys. Res.* **110**, B12104, doi:10.1029/2005/JB003630 (2005).
4. Detrick, R. S. *et al.* Multi-channel seismic imaging of a crustal magma chamber along the East Pacific Rise. *Nature* **326**, 35–41 (1987).
5. Kent, G. M. *et al.* Evidence from three-dimensional reflectivity images for enhanced melt supply beneath mid-ocean-ridge discontinuities. *Nature* **406**, 614–618 (2000).
6. Detrick, R. S., Mutter, J. C., Buhl, P. & Kim, I. I. No evidence from multichannel reflection data for a crustal magma chamber in the MARK area on the Mid-Atlantic Ridge. *Nature* **347**, 61–64 (1990).

7. Calvert, A. J. Backscattered coherence noise and seismic reflection imaging of the oceanic crust: An example from the rift valley of Mid-Atlantic Ridge at 23°N. *J. Geophys. Res.* **102**, 5119–5133 (1997).
8. Navin, D. A., Peirce, C. & Sinha, M. C. The RAMESSES experiment—II. Evidence for accumulated melt beneath a slow spreading ridge from wide-angle refraction and multichannel reflection profiles. *Geophys. J. Int.* **135**, 746–772 (1998).
9. Cannat, M. *et al.* Mid-Atlantic ridge—Azores hotspot interactions: Along-axis migration of a hotspot-derived magmatic pulse 14 to 4 ma ago. *Earth Planet. Sci. Lett.* **173**, 257–269 (1999).
10. Dosso, L. *et al.* The age and distribution of mantle heterogeneity along the Mid-Atlantic Ridge (31–41°N). *Earth Planet. Sci. Lett.* **170**, 269–286 (1999).
11. Schilling, J.-G. Fluxes and excess temperatures of mantle plumes inferred from their interaction with migrating mid-ocean ridges. *Nature* **352**, 397–403 (1991).
12. Detrick, R. S., Needham, H. D. & Renard, V. Gravity anomalies and crustal thickness variations along the Mid-Atlantic Ridge between 33°N and 40°N. *J. Geophys. Res.* **100**, 3767–3787 (1995).
13. Dziak, R. P. *et al.* Evidence of a recent magma dike intrusion at the slow spreading Lucky Strike segment, Mid-Atlantic Ridge. *J. Geophys. Res.* **109**, B12102, doi:10.1029/2004JB003141 (2004).
14. Escartin, J., Cannat, M., Pouliquen, G., Rabain, A. & Lin, J. Crustal thickness of V-shaped ridges south of the Azores; interaction of the Mid-Atlantic Ridge (36°–39° N) and the Azores hotspot. *J. Geophys. Res.* **106**, 21719–21735 (2001).
15. Humphris, S. E., Fornari, D. J., Scheirer, D. S., German, C. R. & Parson, L. M. Geotectonic setting of hydrothermal activity on the summit of Lucky Strike seamount (37°17' N, Mid-Atlantic Ridge). *Geochim. Geophys. Geosyst.* **3**, doi:10.1029/2001GC000284 (2002).
16. Langmuir, C. *et al.* Hydrothermal vents near a mantle hotspot: the Lucky Strike vent field at 37°N on the Mid-Atlantic Ridge. *Earth Planet. Sci. Lett.* **148**, 69–91 (1997).
17. Desbruyères, D. *et al.* Variations in deep-sea hydrothermal vent communities on the Mid-Atlantic Ridge near the Azores plateau. *Deep-Sea Res.* **48**, 1325–1346 (2001).
18. Thatcher, W. & Hill, D. P. A simple model for the fault-generated morphology of slow-spreading mid-ocean ridges. *J. Geophys. Res.* **100**, 561–570 (1995).
19. Calvert, A. J. Seismic evidence for a magma chamber beneath the slow-spreading Mid-Atlantic Ridge. *Nature* **377**, 410–414 (1995).
20. Canales, J. P., Collins, J. A., Escartin, J. & Detrick, R. S. Seismic structure across the rift valley of the Mid-Atlantic ridge at 23°20' N (MARK area): Implications for crustal accretion processes at slow-spreading ridges. *J. Geophys. Res.* **105**, 28411–28425 (2000).
21. Toomey, D. R., Purdy, G. M. & Solomon, S. C. Micro-earthquakes beneath the median valley of the Mid-Atlantic ridge near 23°N: tomography and tectonics. *J. Geophys. Res.* **93**, 9093–9112 (1988).
22. Chen, Y. & Morgan, W. J. Rift valley/no rift valley transition at mid-ocean ridges. *J. Geophys. Res.* **95**, 17571–17581 (1990).
23. MacLennan, J. A., Hulme, T. & Singh, S. C. Thermal models of oceanic crustal accretion: Linking geophysical, geological and petrological observations. *Geochim. Geophys. Geosyst.* **5**, doi:10.1029/2003GC000605 (2003).
24. Smith, D. K. & Cann, J. R. Constructing the upper crust of the Mid-Atlantic Ridge; a reinterpretation based on the Puna Ridge, Kilauea volcano. *J. Geophys. Res.* **104**, 25379–25399 (1999).
25. Meurer, W. P. & Gee, J. S. Evidence for the protracted construction of slow-spread oceanic crust by small magmatic injections. *Earth Planet. Sci. Lett.* **201**, 45–55 (2002).
26. Cannat, M. How thick is the magmatic crust at slow spreading oceanic ridges? *J. Geophys. Res.* **101**, 2847–2857 (1996).
27. Singh, S. C., Kent, G. M., Collier, J. S., Harding, A. J. & Orcutt, J. A. Melt to mush variations in crustal magma properties along the ridge crest at the southern East Pacific Rise. *Nature* **394**, 874–878 (1998).
28. Wilcock, W. S. D. & Delaney, J. R. Mid-ocean ridge sulphide deposits: Evidence for heat extraction from magma chambers or cracking fronts? *Earth Planet. Sci. Lett.* **145**, 49–64 (1996).
29. Dziak, R. P. & Fox, C. G. Long-term seismicity and ground deformation at Axial Volcano, Juan de Fuca Ridge. *Geophys. Res. Lett.* **26**, 3641–3644 (1999).
30. Von Damm, K. L. in *Seafloor Hydrothermal Systems* (eds Humphris, S. E. *et al.*) 222–247 (Geophysical Monograph 91, American Geophysical Union (AGU), Washington DC, 1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the captain (J.-R. Glehen) and the crew of the RV *l'Atalante* for providing support during the SISMOMAR cruise, and the seismic team of GENAVIR for acquiring the seismic data. The INSU MOMAR programme funded the acquisition of SISMOMAR project. This is an Institut de Physique du Globe de Paris contribution.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.C.S. (singh@ipgg.jussieu.fr).

Evidence that mechanisms of fin development evolved in the midline of early vertebrates

Renata Freitas¹, GuangJun Zhang¹ & Martin J. Cohn^{1,2}

The origin of paired appendages was a major evolutionary innovation for vertebrates, marking the first step towards fin- (and later limb-) driven locomotion. The earliest vertebrate fossils lack paired fins but have well-developed median fins^{1,2}, suggesting that the mechanisms of fin development were assembled first in the midline. Here we show that shark median fin development involves the same genetic programs that operate in paired appendages. Using molecular markers for different cell types, we show that median fins arise predominantly from somitic (paraxial) mesoderm, whereas paired appendages develop from lateral plate mesoderm. Expression of *Hoxd* and *Tbx18* genes, which specify paired limb positions^{3,4}, also delineates the positions of median fins. Proximodistal development of median fins occurs beneath an apical ectodermal ridge, the structure that controls outgrowth of paired appendages⁵⁻⁷. Each median fin bud then acquires an anteroposteriorly-nested pattern of *Hoxd* expression similar to that which establishes skeletal polarity in limbs^{8,9}. Thus, despite their different embryonic origins, paired and median fins utilize a common suite of developmental mechanisms. We extended our analysis to lampreys, which diverged from the lineage leading to gnathostomes before the origin of paired appendages^{2,10}, and show that their median fins also develop from somites and express orthologous *Hox* and *Tbx* genes. Together these results suggest that the molecular mechanisms for fin development originated in somitic mesoderm of early vertebrates, and that the origin of paired appendages was associated with re-deployment of these mechanisms to lateral plate mesoderm.

Outgrowth of paired fins and limbs is maintained by the apical ectodermal ridge (AER) at the distal margin of the buds^{3,6}, and members of the Fgf family synergistically mediate its signalling activity^{7,11}. In catsharks, median fins develop from a continuous finfold extending along the dorsal and ventral midlines (Supplementary Fig. 1). Outgrowth of the median finfold occurs beneath an AER-like structure that produces Fgf8 and Dlx proteins (Supplementary Fig. 2). The AER then becomes an apical ectodermal fold (AEF), as in the paired fins of teleosts⁶. The similar embryology of median and paired fins raised the possibility that a common set of mechanisms regulates their development, but their anatomical positions suggested distinctive embryonic origins. Transplantation experiments in amphibians have led to the idea that median finfolds are neural crest derived, and the zebrafish caudal fin was shown to originate, at least in part, from trunk neural crest^{12,13}. Recent fate-mapping studies, however, demonstrated that somitic mesoderm contributes to amphibian median finfold development¹⁴. We therefore set out to determine the embryonic origin of catshark median fins.

Studies in several model systems have shown that *Foxc2* and *Zic1* are expressed in the sclerotome, and that they remain in these cells as they migrate dorsally around the neural tube to form neural arches and spinous processes^{15,16}. Neither of these genes is expressed by

migratory trunk neural crest or differentiated myotome^{15,16}, making them suitable for distinguishing sclerotomal cells during median fin development. We cloned and examined the expression of catshark *Foxc2* and *Zic1* and found that, as in tetrapods, both are expressed throughout the sclerotome (Fig. 1a, b). During median fin development, their expression domains extend dorsally and ventrally into the median finfolds (Fig. 1a, b; Supplementary Figs 3a, b and 4a, b). We also examined *Scleraxis* (sclerotome-related helix-loop-helix type transcription factor), a marker of the sclerotomal sub-compartment (syndetome) that forms axial tendons in chick and mouse^{17,18}. Catshark *Scleraxis* marks a similar subset of the sclerotome and, like *Foxc2* and *Zic1*, its expression domain extends into the median finfolds (Fig. 1c; Supplementary Figs 3c and 4c). Strong expression of all three sclerotomal markers persisted during differentiation of the fin radials and neural arches.

To determine whether cells from the dermomyotome and neural crest also participate in median fin development, we examined *Pax7*, a marker of these cell types in other vertebrate embryos¹⁹. *Pax7* was initially expressed in the catshark dermomyotome and dorsal neural tube, but *Pax7*-expressing cells were not detected in the median fin before stage 31 (Fig. 1d; Supplementary Figs 3d and 4d). *Pax7* expression then extended from the dermomyotome into the median fins, in the muscle projections lateral to the developing skeleton (Fig. 1d; Supplementary Fig. 3d). Immunolocalization of Zn12, a neural crest marker²⁰, revealed that a limited number of neural crest cells also invaded these fins, but most of the mesenchyme was negative for this marker (Fig. 1e; Supplementary Fig. 3e). By stage 31, Zn12 had localized predominantly to the space within the AEF, where dermal rays develop, and subjacent to the distal ectoderm (Fig. 1e; Supplementary Fig. 3e). Together our results suggest that the bulk of the median fin mesenchyme is derived from sclerotome, although cells from dermomyotome and neural crest also contribute to median fin development (Fig. 1f; Supplementary Fig. 3f). If technical challenges can be overcome, cell labelling in shark embryos will further address the contributions of these cell types.

During limb development, lateral plate mesoderm is regionalized into limb-forming and non-limb-forming domains by differential expression of *Hox* and *Tbx* genes^{3,4}. We investigated whether anteroposterior regionalization of the median finfold into dorsal, anal and caudal regions involves similar mechanisms. In catsharks, median fins lie posterior to the cloaca, suggesting that, if *Hox* genes are involved in their development, then the most likely candidates would be *AbdB*-related *Hox9-Hox13* genes. Therefore, we cloned 5' *Hoxd* genes from catsharks and examined their expression during median fin development (Fig. 2 and Supplementary Figs 5 and 6). Prior to the extension of sclerotome towards the dorsal and ventral finfolds, we observed collinear expression of *Hoxd9*, *Hoxd10*, *Hoxd12* and *Hoxd13* in the somitic mesoderm (Supplementary Fig. 6). The *Hoxd9* domain extended anterior to the cloaca, marking the region in which median fin outgrowth was

¹Department of Zoology, ²Department of Anatomy and Cell Biology, University of Florida, PO Box 118525, Gainesville, Florida 32611, USA.

maintained (Supplementary Figs 1 and 6). For *Hoxd9* and *Hoxd10*, we observed different anterior boundaries in the neural tube, paraxial, intermediate and lateral plate mesoderm (Supplementary Fig. 6). The mesenchymal component of the finfold had developed by stage 25, and *Hoxd* genes were expressed in an anteroposteriorly-nested pattern along the dorsal and ventral finfolds, with specific combinations characterizing first dorsal, second dorsal and anal fin levels (Fig. 2a, b). The first dorsal fin region was characterized by expression of *Hoxd9* and *Hoxd10*, whereas the second dorsal and anal regions were distinguished by additional expression of *Hoxd12* (Fig. 2a, b). *Hoxd13* remained confined to the caudal fin region (Fig. 2a, b).

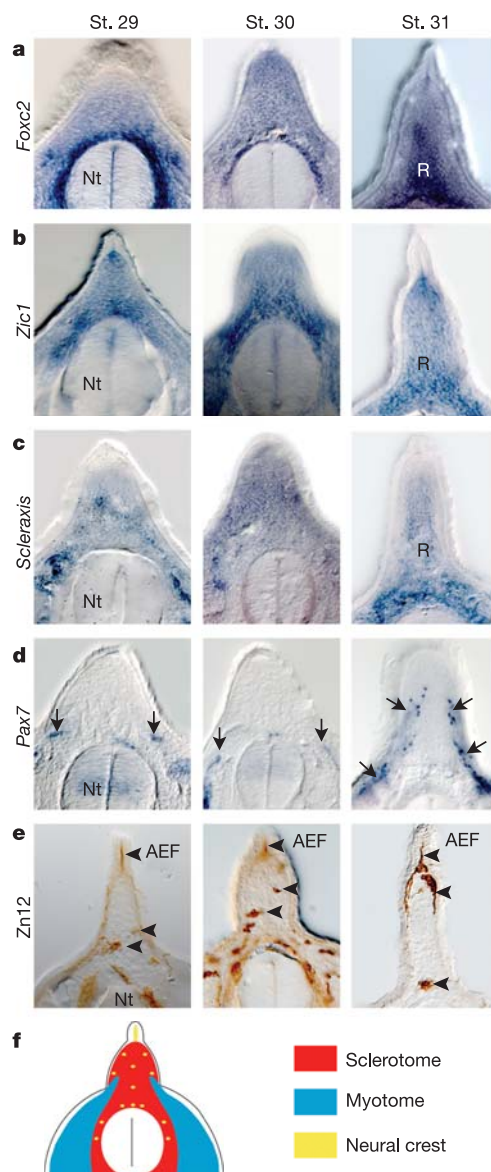


Figure 1 | Developmental origin of catshark median fins. Transverse sections through first dorsal fins; dorsal is to top. Developmental stage (St.) indicated at top. **a–c**, Expression of *Foxc2* (**a**), *Zic1* (**b**) and *Scleraxis* (**c**) in sclerotome surrounding the neural tube (Nt), and in dorsal midline mesenchyme invading the median fins. At stage 31, the strongest expression of *Foxc2* and *Scleraxis* is detected in the developing fin radials (R). **d**, *Pax7* expression in the dorsal lip of the dermomyotome (arrows, stages 29 and 30), and later in myotomal projections invading the median fins (arrows, stage 31). **e**, Isolated Zn12-positive neural crest cells (arrowheads) in median fin mesenchyme and within the apical ectodermal fold (AEF). **f**, Schematic summary of the cellular contributions to the dorsal median fins.

As individual fins emerged from the finfold, *Hoxd* gene expression persisted in the developing fins but was downregulated in the adjacent somites (Fig. 2c). In the first dorsal fin, *Hoxd9* and *Hoxd10* expression was maintained, and *Hoxd12* and *Hoxd13* were activated sequentially (Figs 2c and 3a). The second dorsal and anal fins expressed *Hoxd10*, *Hoxd12* and subsequently *Hoxd13*, but *Hoxd9* expression had shifted anteriorly out of these fins by stage 30 (Fig. 2c). These patterns are consistent with the hypothesis that combinatorial expression of *Hoxd* genes may establish a molecular map for median fin position and identity²¹.

It is unlikely that *Hox* genes act alone to specify fin and limb position. Recent work has implicated *Tbx18* in defining anterior boundaries of forelimbs and somites in chick embryos⁴. To examine whether this gene may also relate to boundary formation in median

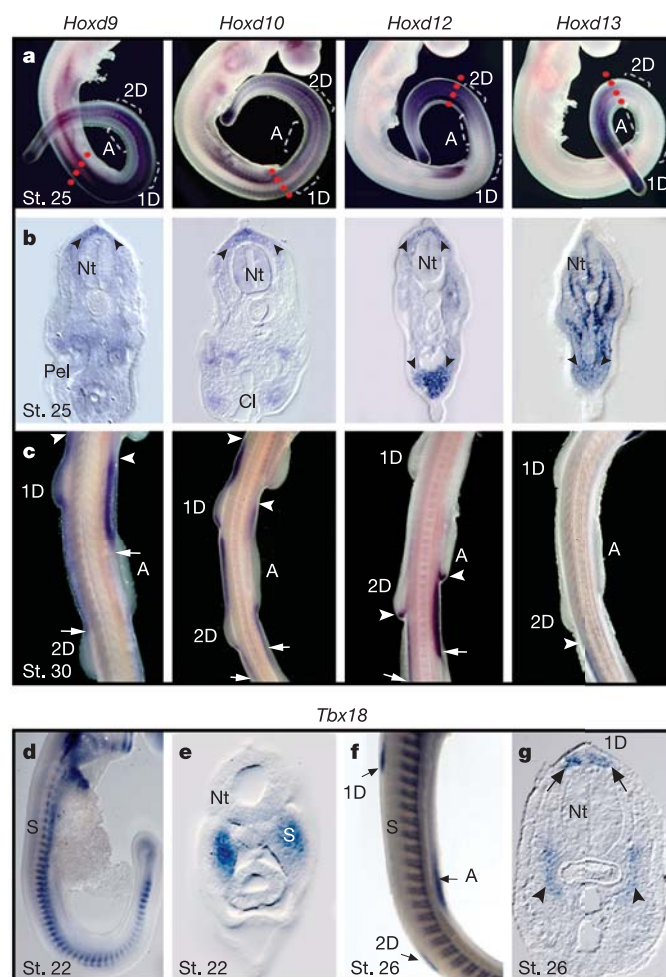


Figure 2 | Regionalized expression of *Hoxd* genes and *Tbx18* along the median finfold of catsharks. Developmental stage indicated in lower left corners. Dorsal is to top in sections and left in whole-mounts. **a**, *Hoxd* gene expression in stage 25 embryos. Brackets indicate regions of prospective first dorsal (1D), second dorsal (2D) and anal (A) fins. Red dotted lines indicate anterior expression boundaries in the median finfold, as verified by histological sections. **b**, Transverse sections immediately posterior to dotted lines in **a** showing *Hoxd* gene expression in dorsal and ventral midline mesenchyme beneath the AEF (arrowheads). Pel, pelvic fin; Cl, cloaca; Nt, neural tube. **c**, *Hoxd* gene expression during emergence of individual median fins at stage 30. Arrowheads mark anterior and arrows mark posterior boundaries of *Hoxd* expression in dorsal and ventral finfolds. **d, e**, Expression of *Tbx18* in somites (S) in whole-mount (**d**) and transverse section (**e**). **f**, *Tbx18* expression becomes detectable in presumptive dorsal and anal fins at stage 26 (arrows). **g**, Transverse section showing *Tbx18* expression in first dorsal fin (arrows) and in sclerotome (arrowheads).

fin, we cloned and examined expression of a catshark *Tbx18* orthologue. *Tbx18* was first expressed in the anterior region of each somite (Fig. 2d, e). During finfold outgrowth, we detected *Tbx18* in three discrete domains that delineated the prospective first dorsal, second dorsal and the anal fins (Fig. 2f, g), resembling the pattern observed in chick limbs⁴. Given the function of *Tbx18* in specifying limb position in lateral plate mesoderm, we suggest that *Tbx18* may also participate in specification of median fin position within the finfold, further extending the parallel between median and paired fin development.

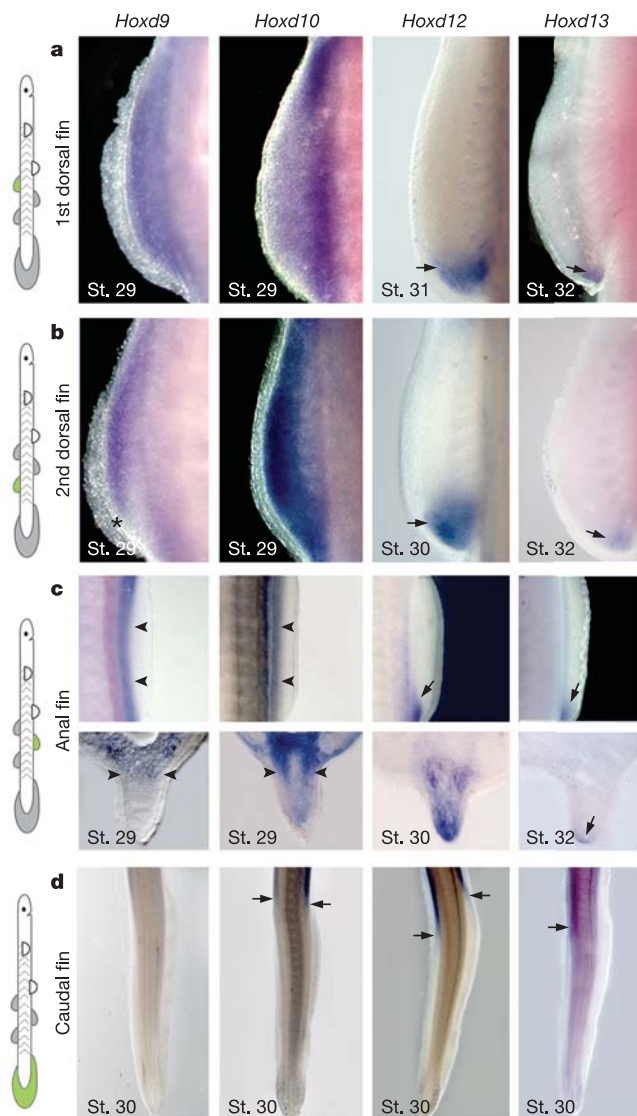


Figure 3 | Anteroposterior nesting of *Hoxd* gene expression in catshark median fin buds. Anterior is to top and dorsal is to left in whole-mounts; dorsal is to top in sections. Stages of development indicated in lower left corners. Diagram to left shows location of fins (green) depicted in adjacent panels. **a**, *Hoxd* expression in first dorsal fins. Arrows point to *Hoxd12* and *Hoxd13* expression at the posterior margin of fin. **b**, *Hoxd* expression in second dorsal fins. Asterisk marks the posterior boundary of *Hoxd9* expression that has started to shift anteriorly. Arrows point to expression of *Hoxd12* and *Hoxd13* at the posterior margin of the fin. **c**, Upper panels show *Hoxd* expression in anal fins, and lower panels show transverse sections through these fins. Arrowheads indicate *Hoxd9* and *Hoxd10* expression along the proximal region of fin. Arrows mark posterior expression of *Hoxd12* and *Hoxd13*. Note temporal and spatial collinearity of *Hoxd* gene expression within the dorsal (**a**, **b**) and anal (**c**) fins. **d**, Caudal fins lacking *Hoxd* expression at stage 30; arrows indicate posterior limits of *Hoxd* expression in the pre-caudal finfolds.

Dorsal and anal fin skeletons are polarized along their antero-posterior axes (Supplementary Fig. 1d, e). In paired limb buds, *Hoxd* genes are expressed along the anteroposterior axis in a spatial and temporally collinear pattern that determines the polarity of the skeleton^{8,9}. We therefore explored whether *Hoxd* expression in each median fin follows the patterns observed in paired limbs. In the first dorsal fin, *Hoxd9* and *Hoxd10* were expressed broadly from the onset of budding (Fig. 3a). *Hoxd12* became detectable posteriorly by stage 31, and *Hoxd13* expression was observed nested within the *Hoxd12* domain one stage later (Fig. 3a). Similar collinear patterns were observed in the second dorsal and anal fins (Fig. 3b, c). Thus, shark dorsal and anal fins exhibit the characteristic collinearity of paired appendages^{22,23}. In contrast to this 'appendicular' pattern of expression in dorsal and anal fins, the caudal fin develops in the absence of *Hoxd* gene expression after stage 30 (Fig. 3d). Posterior expression of 5' *Hoxd*, *Hoxb8* and *Hand2* genes establishes the zone of polarizing activity in paired limbs, which patterns the anteroposterior axis via secretion of Sonic hedgehog⁸. Our finding of a conserved relationship between the polarity of *Hoxd* gene expression and the anteroposterior pattern of the fin skeleton suggests that a similar mechanism may operate in median fins.

Interpretation of these results in the context of fin evolution suggests that the fin development program may have originated in paraxial mesodermally-derived median fins before paired fins evolved in lateral plate mesoderm. To test this hypothesis, we

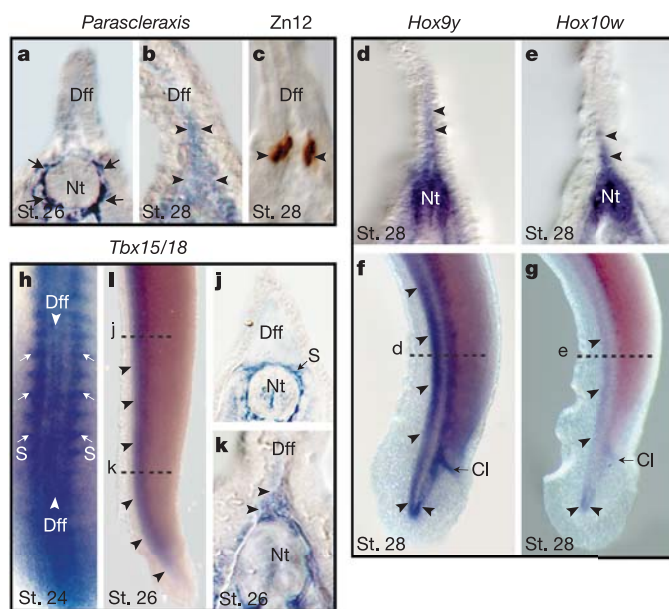


Figure 4 | Lamprey median fin development. Stages of development indicated in lower left corners. **a–e**, **j** and **k** are transverse sections with dorsal to top, **f–i** are whole-mounts with dorsal to left. **a**, Expression of *Parascleraxis* in sclerotome (arrows) adjacent to neural tube (Nt). Dff, dorsal finfold. **b**, *Parascleraxis* expression has expanded into the dorsal finfold (arrowheads). **c**, Zn12 staining in two clusters of cells at the base of the dorsal finfold (arrowheads). **d**, **e**, Expression of *Hox9y* (**d**) and *Hox10w* (**e**) in the dorsal finfold mesenchyme (arrowheads). Note that the fin tissue distal to the expression domains is ectodermal. **f**, **g**, Expression of *Hox9y* (**f**) and *Hox10w* (**g**) in the finfold (arrowheads). Dashed lines mark the approximate planes of sections shown in **d** and **e**. Cl, cloaca. **h**, Expression of *Tbx15/18* in the anterior part of each somite (S; arrows) and in the dorsal finfold (arrowheads) at stage 24. **i**, Expression of *Tbx15/18* in the median finfold at stage 26 (arrowheads). Dashed lines mark approximate planes of sections shown in **j** and **k**. **j**, **k**, Transverse sections taken anterior (**j**) and posterior (**k**) to the expression boundary of *Tbx15/18* in the median finfold. Note that somitic expression extends anterior to the finfold domain (**k**).

extended our analysis to lampreys, which exhibit the plesiomorphic condition of median fins in the absence of paired appendages^{2,10}. During lamprey embryogenesis, a single ectodermal median finfold develops along the entire trunk. Proximodistal expansion of finfold mesenchyme occurs predominantly in the posterior region, where median fins differentiate during metamorphosis²⁴. To determine whether median fins of lampreys and sharks have the same embryonic origin, we isolated and characterized the expression of a lamprey *Scleraxis* orthologue. Phylogenetic analysis of our full-length clone placed it as the sister to the gnathostome *Scleraxis*/Paraxis clade, and therefore we designated it *Parascleraxis* (Supplementary Fig. 7). *Parascleraxis* expression was detected at stage 26 in sclerotomal cells adjacent to the neural tube but not in dermomyotome (Fig. 4a). By stage 28, the *Parascleraxis* domain extended into the median finfold, consistent with a sclerotomal contribution to lamprey median fins (Fig. 4b). Restriction of *Parascleraxis* to the lamprey sclerotome also suggests that the dermomyotomal domain of *Paraxis* in gnathostomes may be a novel site of expression that was acquired after duplication of the ancestral *Parascleraxis* gene gave rise to *Paraxis* and *Scleraxis*. We then stained lamprey embryos for Zn12, as previous workers reported migration of neural crest cells into the lamprey median finfold^{25,26}. Zn12 signal was detected in two clusters of sub-epidermal cells at the base of the finfold, and later in a narrow column of cells at its distal tip, but the bulk of median finfold mesenchyme was negative for Zn12 (Fig. 4c and data not shown). These data suggest that, as in sharks, the sclerotome and a limited number of neural crest cells give rise to the median fin mesenchyme of lampreys.

We next asked whether anteroposterior boundaries of median fins in lampreys and sharks are specified by an evolutionarily conserved mechanism involving 5' *Hox* and *Tbx18* orthologues. We analysed the expression of lamprey *Hox9y* and *Hox10w* (ref. 27), and detected expression of both genes in the median finfold mesenchyme (Fig. 4d, e). The anterior boundary of *Hox9y* expression in the dorsal finfold and adjacent somites extended anterior to the *Hox10w* domain and delineated the region in which dorsal fin outgrowth is sustained during larval development (Fig. 4f, g). We also screened for a lamprey *Tbx18* orthologue and isolated a complementary DNA fragment that our phylogenetic analyses joined to the base of the gnathostome *Tbx15/18* clade (Supplementary Fig. 8). Lamprey *Tbx15/18* was expressed anteriorly in each somite and in the median finfold at stage 24 (Fig. 4h). Expression in the median finfold mesenchyme remained posterior, in the fin-forming region, whereas the somitic expression extended along the entire trunk at stage 26 (Fig. 4i–k).

Conservation of the embryonic origin and the patterns of *Hox* and *Tbx* gene expression in shark and lamprey median fins suggests that the molecular mechanisms of fin development evolved in the midline before the origin of paired fins. Our finding that median fin mesenchyme arises predominantly from somites suggests that these cells may acquire their positional identities, in the form of *Hox* and *Tbx* expression, during regionalization of paraxial mesoderm. We suggest that the origin of paired appendages from lateral plate mesoderm involved re-deployment of mechanisms that were originally restricted to paraxial mesoderm, where they regulated development of cartilage (*Hox9–Hox13*, *Tbx18*), muscle (*Pax7*, *Paraxis*) and tendon (*Scleraxis*) in the axial skeleton and median fins. Reports of *Msx* and *Dlx* expression in paired and median fins of zebrafish^{28,29} may reflect additional evolutionary signatures of this co-option. It is possible that the mechanisms of fin and limb development were established in median finfolds even before the origin of vertebrates. Analysis of median finfold development in cephalochordates will further test the hypothesis that these mechanisms emerged early in chordate evolution.

METHODS

Isolation of catshark and lamprey genes. Degenerate polymerase chain reaction with reverse transcription (RT–PCR) was performed to amplify catshark

fragments of 5' *Hoxd* genes (*Hoxd9*, 387 base pairs, bp; *Hoxd10*, 813 bp; *Hoxd12*, 561 bp; *Hoxd13*, 579 bp), *Tbx18* (534 bp), *Zic1* (204 bp), *Foxc2* (267 bp) and *Pax7* (294 bp) using cDNA from a stage 28 *Scyliorhinus canicula*. A full-length copy of catshark *Scleraxis* (1,413 bp) was obtained by RT–PCR and 5' rapid amplification of cloned ends (RACE). Lamprey genes (*Tbx15/18* and *Parascleraxis*) were isolated by RT–PCR from a *Petromyzon marinus* cDNA library (a gift from J. Langeland) using Advantage GC–PCR Kit (Clontech). The amplified fragments were cloned into pGEM-T Easy Vector (Promega) or pDrive Cloning Vector (Qiagen). Orthology of the cloned sequences was determined by protein alignment comparisons followed by maximum-likelihood and neighbour-joining methods.

Whole-mount *in situ* hybridization and immunohistochemistry. These were performed as described previously³⁰. Antibodies against Fgf8 (Santa Cruz Biotechnology Inc), Distal-less (Dll/Dlx; kindly supplied by G. Boekhoff-Falk) and Zn12 (Developmental Studies Hybridoma Bank) were diluted to working concentrations of 1:100, 1:70 and 1:5 respectively. Peroxidase-conjugated secondary antibodies (DAKO) were diluted to 1:500 in phosphate buffered saline (PBS) with 1% Triton and 1% serum. Following whole-mount *in situ* hybridization or immunohistochemistry, the specimens were equilibrated in graded sucrose in PBS (15% and 30%) at 4 °C and graded gelatine in PBS (20% gelatine in 30% sucrose and 20% gelatine) at 50 °C. Embryos were then mounted in Tissue-Tek OCT (Sakura Finetek) and cryosectioned at 20–35 µm.

Received 26 July 2005; accepted 19 June 2006.

Published online 26 July 2006.

1. Zhang, X. G. & Hou, X. G. Evidence for a single median fin-fold and tail in the Lower Cambrian vertebrate, *Haikouichthys ercaicunensis*. *J. Evol. Biol.* **17**, 1162–1166 (2004).
2. Coates, M. I. The origin of vertebrate limbs. *Development (Suppl.)*, 169–182 (1994).
3. Cohn, M. J. *et al.* *Hox9* genes and vertebrate limb specification. *Nature* **387**, 97–101 (1997).
4. Tanaka, M. & Tickle, C. *Tbx18* and boundary formation in chick somite and wing development. *Dev. Biol.* **268**, 470–480 (2004).
5. Saunders, J. W. The proximo-distal sequence of origin of parts of the chick wing and the role of the ectoderm. *J. Exp. Zool.* **108**, 363–403 (1948).
6. Grandel, H. & Schulte-Merker, S. The development of the paired fins in the zebrafish (*Danio rerio*). *Mech. Dev.* **79**, 99–120 (1998).
7. Grandel, H., Draper, B. W. & Schulte-Merker, S. *dackel* acts in the ectoderm of the zebrafish pectoral fin bud to maintain AER signaling. *Development* **127**, 4169–4178 (2000).
8. Zakany, J., Kmita, M. & Duboule, D. A dual role for *Hox* genes in limb anterior-posterior asymmetry. *Science* **304**, 1669–1672 (2004).
9. Turchini, B. & Duboule, D. Control of *Hoxd* genes collinearity during early limb development. *Dev. Cell* **10**, 93–103 (2006).
10. Donoghue, P. C. J., Forey, P. L. & Aldridge, R. J. Conodont affinity and chordate phylogeny. *Biol. Rev. Camb. Philos. Soc.* **75**, 191–251 (2000).
11. Sun, X., Mariani, F. V. & Martin, G. R. Functions of FGF signalling from the apical ectodermal ridge in limb development. *Nature* **418**, 501–508 (2002).
12. Tucker, A. S. & Slack, J. M. Independent induction and formation of the dorsal and ventral fins in *Xenopus laevis*. *Dev. Dyn.* **230**, 461–467 (2004).
13. Smith, M., Hickman, A., Amanze, D., Lumsden, A. & Thorogood, P. Trunk neural crest origin of caudal fin mesenchyme in the zebrafish *Brachydanio rerio*. *Proc. R. Soc. Lond. B* **256**, 137–145 (1994).
14. Sobkow, L., Epperlein, H. H., Herklotz, S., Straube, W. L. & Tanaka, E. M. A germline GFP transgenic axolotl and its use to track cell fate: dual origin of the fin mesenchyme during development and the fate of blood cells during regeneration. *Dev. Biol.* **290**, 386–397 (2006).
15. Furumoto, T. A. *et al.* Notochord-dependent expression of MFH1 and PAX1 cooperates to maintain the proliferation of sclerotome cells during the vertebral column development. *Dev. Biol.* **210**, 15–29 (1999).
16. Sun Rhodes, L. S. & Merzdorf, C. S. The *zic1* gene is expressed in chick somites but not in migratory neural crest. *Gene Expr. Patterns* **6**, 539–545 (2006).
17. Brent, A. E., Schweitzer, R. & Tabin, C. J. A somitic compartment of tendon progenitors. *Cell* **113**, 235–248 (2003).
18. Brent, A. E. & Tabin, C. J. FGF acts directly on the somitic tendon progenitors through the Ets transcription factors *Pea3* and *Erm* to regulate scleraxis expression. *Development* **131**, 3885–3896 (2004).
19. Lacosta, A. M., Muniesa, P., Ruberte, J., Sarasa, M. & Dominguez, L. Novel expression patterns of Pax3/Pax7 in early trunk neural crest and its melanocyte and non-melanocyte lineages in amniote embryos. *Pigment Cell Res.* **18**, 243–251 (2005).
20. Trevarrow, B., Marks, D. L. & Kimmel, C. B. Organization of hindbrain segments in the zebrafish embryo. *Neuron* **4**, 669–679 (1990).
21. Mabee, P. M., Crotwell, P. L., Bird, N. C. & Burke, A. C. Evolution of median fin modules in the axial skeleton of fishes. *J. Exp. Zool.* **294**, 77–90 (2002).
22. Sordino, P., Van der Hoeven, F. & Duboule, D. *Hox* gene expression in teleost fins and the origin of vertebrate digits. *Nature* **375**, 678–681 (1995).

23. Nelson, C. E. *et al.* Analysis of Hox gene expression in the chick limb bud. *Development* **122**, 1449–1466 (1996).
24. Richardson, M. K. & Wright, G. M. Developmental transformations in a normal series of embryos of the sea lamprey *Petromyzon marinus* (Linnaeus). *J. Morphol.* **257**, 348–363 (2003).
25. Hirata, M., Ito, K. & Tsunekuni, K. Migration and colonization patterns of HNK-1-immunoreactive neural crest cells in lamprey and swordtail embryos. *Zool. Sci.* **14**, 305–312 (1997).
26. McCauley, D. W. & Bronner-Fraser, M. Neural crest contributions to the lamprey head. *Development* **130**, 2317–2327 (2003).
27. Force, A., Amores, A. & Postlethwait, J. H. Hox cluster organization in the jawless vertebrate *Petromyzon marinus*. *J. Exp. Zool.* **294**, 30–46 (2002).
28. Akimenko, M. A., Ekker, M., Wegner, J., Lin, W. & Westerfield, M. Combinatorial expression of three zebrafish genes related to *distal-less*: part of a homeobox gene code for the head. *J. Neurosci.* **14**, 3475–3486 (1994).
29. Akimenko, M. A., Johnson, S. L., Westerfield, M. & Ekker, M. Differential induction of four *msx* homeobox genes during fin development and regeneration in zebrafish. *Development* **121**, 347–357 (1995).
30. Freitas, R. & Cohn, M. J. Analysis of *EphA4* in the lesser spotted catshark identifies a primitive gnathostome expression pattern and reveals co-option during evolution of shark-specific morphology. *Dev. Genes Evol.* **214**, 466–472 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Burke, P. Mabey, P. Crotwell and B. Shockey for commenting on the manuscript, A. Graham for sharing reagents, and L. Page and G. Weddle for assistance with lamprey collection. R. Freitas is a PhD student of the GABBA Program (ICBAS, Univ. Oporto) and was supported by a fellowship from FCT, Praxis XXI.

Author contributions R.F. performed and designed (with M.J.C.) the reported studies. G.Z. performed part of the gene cloning and phylogenetic analyses. M.J.C. supervised the research project, and assisted in the experimental design. R.F. and M.J.C. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Sequences for *Foxc2*, *Zic1*, *Scleraxis*, *Pax7*, *Hoxd9*, *Hoxd10*, *Hoxd12*, *Hoxd13* and *Tbx18* from *S. canicula*, and *Parascleraxis* and *Tbx15/18* from *P. marinus*, are deposited in GenBank under accession numbers DQ659101–DQ659111. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.J.C. (cohn@zoo.ufl.edu).

LETTERS

The emergence of geometric order in proliferating metazoan epithelia

Matthew C. Gibson^{1*†}, Ankit B. Patel^{2*}, Radhika Nagpal² & Norbert Perrimon¹

The predominantly hexagonal cell pattern of simple epithelia was noted in the earliest microscopic analyses of animal tissues¹, a topology commonly thought to reflect cell sorting into optimally packed honeycomb arrays². Here we use a discrete Markov model validated by time-lapse microscopy and clonal analysis to demonstrate that the distribution of polygonal cell types in epithelia is not a result of cell packing, but rather a direct mathematical consequence of cell proliferation. On the basis of *in vivo* analysis of mitotic cell junction dynamics in *Drosophila* imaginal discs, we mathematically predict the convergence of epithelial topology to a fixed equilibrium distribution of cellular polygons. This distribution is empirically confirmed in tissue samples from vertebrate, arthropod and cnidarian organisms, suggesting that a similar proliferation-dependent cell pattern underlies pattern formation and morphogenesis throughout the metazoa.

From sponges to humans, the organization of cells into epithelial sheets is an essential feature of animal design. Historically, the characteristic 'cobblestone' topology of monolayer epithelia has been presumed to reflect optimal cell packing. Many biological and non-biological systems do indeed form predictable geometric arrays due to a tendency either to minimize surface energy or maximize space filling. Prominent examples include insect retinal cells, non-proliferating epithelia, honeycombs, compressed soap bubbles, and even coins pushed together on a tabletop^{2–5}. However, in contrast with these systems, proliferating epithelia rarely exhibit a honeycomb pattern (Fig. 1a), more often forming irregular polygon arrays due to the effect of cell division^{6–9}. The wing primordium (imaginal disc) of the fruitfly *Drosophila melanogaster*, for example, is an epithelial sheet that grows from ~20 to ~50,000 cells during approximately 4 days of larval development^{10–12}. At the level of the adhesive septate junctions that bind cells to their neighbours¹³, the wing epithelium is a heterogeneous lattice dominated by hexagons, but also featuring significant numbers of four- to nine-sided cells (Fig. 1b). Although recent progress has been made in understanding rearrangements of epithelial topology during morphogenesis^{14–16}, the mechanisms that determine cell topology in proliferating epithelia remain poorly defined.

To understand the dynamic process that generates the heterogeneous cell pattern in the *Drosophila* wing, time-lapse movies were collected using fluorescent proteins that localize to the septate junction (ATPase- α -GFP and neuroglian-GFP^{17,18}). Consistent with a negligible role for cell rearrangement in the determination of topology, large-scale sorting or migration within the epithelium was not observed (Supplementary Movie 1). The only significant cellular movements occurred during mitosis, as initially polygonal prophase cells rounded up and divided into two daughter polygons (Fig. 1c; see also Supplementary Movies 2 and 3). Despite marked dilation of mitotic cells in prophase-metaphase (Fig. 1d) and the subsequent contraction of the cytokinetic furrow (Fig. 1e),

cell-neighbour relationships were stably maintained throughout the cell cycle, attesting to an elastic capacity of the junctional lattice to conserve topology over time. These results indicate that cells tightly adhere to their immediate neighbours, consistent with the well-established formation of contiguous cell lineage clones in *Drosophila* appendage primordia^{19–21}.

To characterize more precisely how cells form new interfaces after division, we used the FLP-out technique²² to activate stochastically Gal4 transcription factor activity in individual cells, thereby marking single cell lineages with expression of a *UAS-GFP* transgene. Small GFP⁺ cell clones were then scored to determine directly the spatial relationship between post-mitotic siblings (Fig. 1f). At least 94% of mitoses resulted in the apposition of daughters across a common septate junctional interface (type I, Fig. 1g; $n = 250$ clones). These observations indicate that cytokinesis leads to a transient bottle-neck conformation (for example, Fig. 1e) that predictably resolves with abscission into two polygonal cells adjoined by a common side (Fig. 1h).

If epithelial cells adhere to their neighbours and do not re-sort, then cell division should be sufficient to account for the heterogeneous topology of monolayer epithelia, including the predominance of hexagons. To test mathematically this hypothesis, we defined six logical conditions: (1) cells are polygons with a minimum of four sides ($n = 2,172$ wing disc cells); (2) cells do not re-sort (Supplementary Movies 1–3); (3) mitotic siblings retain a common junctional interface (for example, Fig. 1f); (4) cells have asynchronous but roughly uniform cell cycle times¹²; (5) cleavage planes always cut a side rather than a vertex of the mother polygon (inferred from the observation that 4-cell junctions are rare and presumably unstable); and (6) mitotic cleavage orientation randomly distributes existing tricellular junctions to both daughter cells.

Using conditions 1–5, the wing epithelium can be formulated as a two-dimensional planar network (graph). In topological terms, each tricellular junction is a vertex, each cell side is an undirected edge, and each apical cell surface is a polygonal face. Let v_t , e_t and f_t denote the number of vertices, edges and faces after t divisions. If we assume that cells divide at a uniform rate, then the number of faces (cells) will double after each round of division. Thus, $f_t = 2f_{t-1}$. Furthermore, because each cell division results in biogenesis of two vertices and three edges (for example, Fig. 1h), the number of vertices at time t is $v_t = v_{t-1} + 2f_{t-1}$ and the number of edges is $e_t = e_{t-1} + 3f_{t-1}$. Because boundary effects become negligible for large t (Supplementary Data), we can approximate the average number of sides per cell (s) at division t :

$$s_t = 2(e_{t-1} + 3f_{t-1})/2f_{t-1} = (s_{t-1}/2) + 3 \quad (1)$$

This recurrence system is solvable for the state of the epithelial network as a function of the initial network at time $t = 0$:

$$s_t = 6 + 2^{-t}(s_0 - 6) \quad (2)$$

¹Department of Genetics and Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA. ²Division of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA. [†]Present address: Stowers Institute for Medical Research, 1000 East 50th Street, Kansas City, Missouri 64110, USA. *These authors contributed equally to this work.

Equation (2) shows that the average number of cell sides exponentially approaches six, consistent with Euler's theorem. This implies that even in epithelia devoid of minimal packing, the system will assume a predominantly hexagonal topology as a consequence of cell

division. This behaviour is independent of cleavage plane orientation, and is instead a result of the formation of tricellular junctions, as previously demonstrated for plant tissues⁷⁻⁹. Importantly, however, an average of six does not necessitate a prevalence (or even existence) of hexagons and so a higher fidelity model is required.

We formulated a more precise model using a discrete Markov chain to capture the stochastic nature of cell proliferation, inspired by mathematical work on random space-dividing topologies²³. We defined the state of a cell, s , as its number of sides where $s > 3$. The relative frequency of s -sided cells in the population was defined as p_s , and the state of the population at generation t as an infinite row vector $\mathbf{p}^{(t)} = [p_4 p_5 p_6 p_7 p_8 p_9 \dots]$. The state dynamics is described by $\mathbf{p}^{(t+1)} = \mathbf{p}^{(t)}PS$, where P and S are probabilistic transition matrices (Box 1). Briefly, the entries P_{ij} represent the probability that an i -sided cell will become j -sided after mitosis. Topological arguments indicate that a cell will gain an average of one new side per cycle due to neighbour divisions, and the matrix S accounts for this effect. Thus, given the distribution of polygonal cell types $\mathbf{p}^{(t)}$, we can compute the new distribution after a single round of division.

Formulating epithelial topology as a Markov chain (Fig. 2a) yields a strong quantitative prediction: that a stable equilibrium distribution of polygons should emerge in proliferating epithelia, irrespective of the

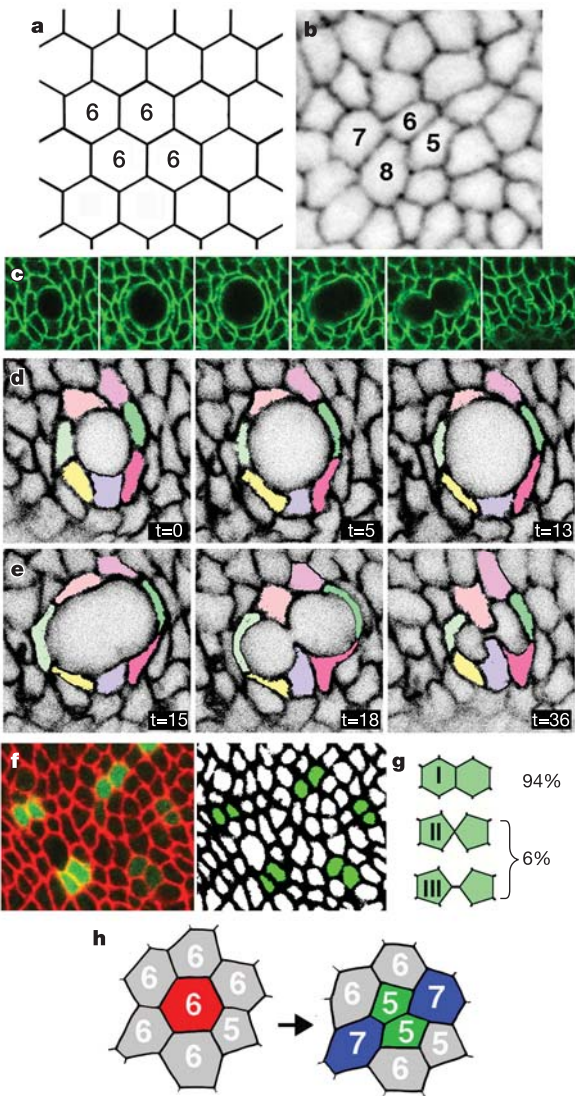


Figure 1 | Mitosis and the *in vivo* dynamics of epithelial topology. **a**, The regular hexagonal array typical of free energy minimization processes is defined by uniformity of cell side length and the formation of tricellular junctions, with each intersecting cell side separated by an equivalent 120° angle. **b**, At the level of the septate junctions (stained here for Discs large (Dlg; black)), cell topology in the wing disc epithelium is highly irregular. **c**, Six successive stages of cell division from a confocal time-lapse recording. Septate junction dynamics, monitored with nrg-GFP (green), show that mitotic cells first round up and then divide at the apical epithelial surface. **d, e**, Greyscale-inverted images from **a** showing conservation of cell contacts throughout cell division. **d**, Dilation of the junctional lattice permits rounding of a seven-sided mitotic cell during stages corresponding to prophase–metaphase. Owing to compression and stretching of the pseudo-coloured neighbours, no cell–neighbour exchanges occur ($n = 18$ dilating cells). Units of t are in minutes. **e**, During stages corresponding to anaphase through cytokinesis, local topology (connectivity between cells) remains unchanged; the mitotic cell approaches abscission surrounded by the same cohort of seven neighbours ($n = 23$ cytokinetic cells). **f**, Two-cell clones marked by heritable expression of GFP (green) imaged at the level of the septate junctions stained with anti-Dlg (red). **g**, In approximately 94% of cell divisions, cytokinesis resolves with formation of a new cell interface, resulting in the type I conformation of mitotic siblings. **h**, Summary diagram of topology changes during cell division.

Box 1 | Derivation of Markov state dynamics

Here we derive the probabilistic transition matrices P and S . The entry P_{ij} is the probability that an i -sided cell divides to produce a j -sided daughter cell. Consider a single cell with s_{t-1} sides (or junctions) at generation $t - 1$. Let the random variable K_t represent the number of junctions distributed to one daughter cell on division at generation t , leaving $s_{t-1} - K_t$ for the other. Because no triangular cells are observed empirically, we assume that each daughter receives at least two junctions from the parent, leaving $s_{t-1} - 4$ junctions to be distributed among the daughters. Assuming that junctions are distributed uniformly at random around the mitotic cell and that cleavage plane orientation is chosen uniformly at random (to bisect the rounded mitotic cell's area), we can model the distribution of these remaining junctions as balls thrown into one of two bins (daughters) with equal probability. Thus, the number of additional parental junctions received by the first daughter is $K_t - 2$, a binomial random variable with parameters $n = s_{t-1} - 4$, $p = \frac{1}{2}$. Finally, each daughter also gains two new junctions as a result of the newly created interface. Therefore, the probability of transition from an i -sided cell to a j -sided daughter is $P_{ij} = \Pr[K_t + 2 = j | s_{t-1} = i] = \text{Comb}(i - 4, j - 4) / 2^{i-4}$, where $\text{Comb}(a, b)$ is the number of ways to choose b objects from a set of a objects. As a consequence, the (un-normalized) entries of P are the coefficients of Pascal's triangle.

Next we derive the 'shift matrix' S , the entries, S_{ij} , of which represent the probability that an i -sided cell will gain sides from dividing neighbours to become j -sided. Thus, S accounts for the effect of neighbour cell divisions on the polygon class of a given cell, an effect that was unaccounted for in previous work²³. On mitosis, a cell adds one side to each of two neighbouring cells. Assuming N cells in an epithelium, this means that $+2N$ sides are added during one round of divisions, resulting in $2N$ cells. Hence, the average number of sides gained per cell is $+2N/2N = +1$. That is, a cell will gain, on average, one side per cycle from dividing neighbours. Thus, the entries of the matrix S are $S_{ij} = 1$ if $j = i + 1$ and zero otherwise. Note that this is a mean-field approximation of the effect of dividing neighbour cells. In reality, some cells will gain no sides and others will gain more than one side.

P	Post-mitotic class							S	After neighbour divisions							
	4	5	6	7	8	9	10		4	5	6	7	8	9	10	
Pre-mitotic class	4	1						Before neighbour divisions	4	0	1					
	5	1	1						5		0	1				
	6	1	2	1					6			0	1			
	7	1	3	3	1				7				0	1		
	8	1	4	6	4	1			8					0	1	
	9	1	5	10	10	5	1		9						0	1
	10						...		10							...

initial conditions (Perron–Frobenius theorem)²⁴. We calculated the exact equilibrium E directly from the matrix $T = PS$ to be approximately 28.9% pentagons, 46.4% hexagons, 20.8% heptagons and lesser frequencies of other polygon types (Fig. 2b; see Methods). The model also predicts that the population of cells should approach this distribution at an exponential rate. Consequently, global topology converges to E in less than eight generations, even for initial conditions where every cell is quadrilateral, hexagonal, or nonagonal (Fig. 2c). Furthermore, we modelled defective interface formation between mitotic siblings, and found that E is robust to small error rates (Supplementary Data).

To validate whether E exists *in vivo*, we determined the actual polygon distribution in the developing *Drosophila* wing. Notably, empirical counts closely matched E for every major polygon class to within a few per cent ($n = 2,172$ cells; Fig. 3a). Only the small percentage of four-sided cells was unaccounted for, an effect attributable to the mean field approximation (Box 1 and Supplementary Data). In addition, the variation between individual imaginal discs was minimal (see error bars for Fig. 3a; see also Supplementary Fig. 2). Our first-order Markov model can therefore explain global topology in the *Drosophila* wing disc epithelium, including the predominance of hexagons, based on the mechanism of cell division alone. A unique advantage of this simple model is its generality, particularly given the conservation of fundamental aspects of adhesion and mitosis in monolayer epithelia. Consistent with this, we found that polygon topology in the tadpole tail epidermis of the frog *Xenopus* ($n = 1,051$ cells) and in the outer epidermis of the freshwater cnidarian *Hydra* ($n = 602$ cells) also closely approached E (Fig. 3a). Although the details of cell division in these organisms remain to be described, we infer that the same equilibrium topology will emerge in most multicellular eukaryotes. Indeed, the reported cell topology in plant epidermis is in remarkably close agreement with our results (25.1% pentagons, 47.4% hexagons and 22.4% heptagons)⁸.

In addition to predicting the equilibrium topology, our model also

provides a quantitative framework for further insights into epithelial organization. For example, the model predicts that each cell in the population must gain an average of one side per cell cycle (Box 1). Confirming this, we found that the average mitotic cell (at the end of the cell cycle) possessed not six (5.94 ± 0.06) but rather seven sides (6.99 ± 0.07 ; Fig. 3b, c; $n = 177$ mitotic figures). This indicates that epithelial cells accumulate additional sides until mitosis, at which time they divide into two daughters of lesser sidedness. Topological equilibrium is therefore achieved as a balance between autonomous and non-autonomous division events.

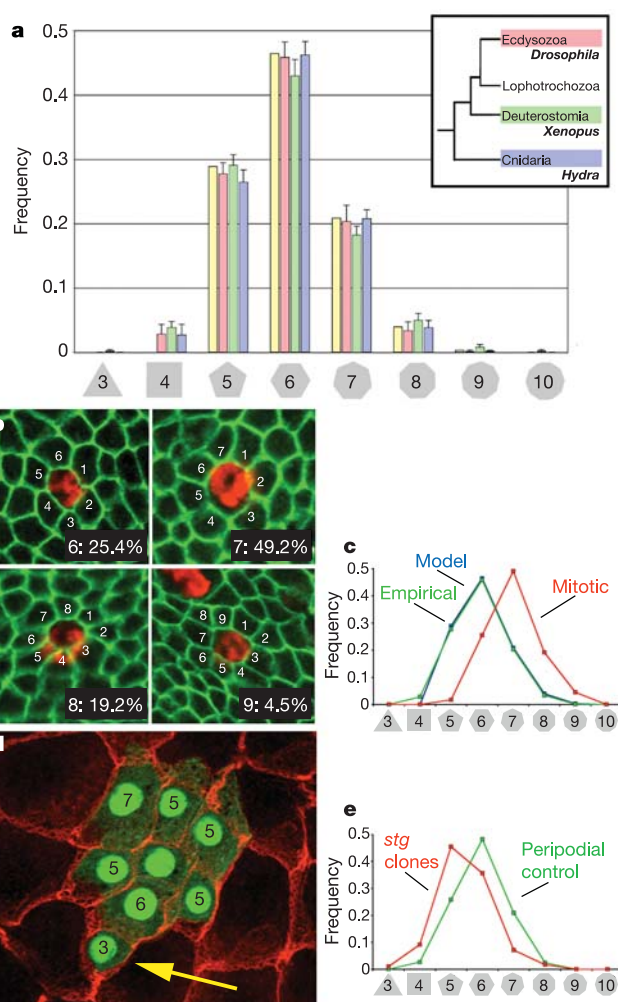


Figure 3 | An emergent topological order in proliferating epithelia. **a**, In close accordance with the theoretical equilibrium topology (yellow), *Drosophila* wing disc (pink, $n = 2,172$ cells), *Xenopus* tail epidermis (green, $n = 1,051$ cells) and *Hydra* epidermis (blue, $n = 602$ cells) all exhibit a similar non-gaussian distribution of epithelial polygons with less than 50% hexagonal cells and high (and asymmetric) percentages of pentagonal and heptagonal cells. Error bars indicate standard deviation between individual samples. The inset indicates relative phylogenetic positions for *Drosophila*, *Xenopus* and *Hydra*²⁶. **b**, Prophase cells (marked with anti-phospho-histone H3; red) stained for cell junctions (Dlg, green) to quantify cell sidedness. Most mitotic cells are seven-sided. **c**, Polygonal cells in the Markov model (predicted, blue) and *Drosophila* wing disc (empirical, green) have an average of six sides. The population of mitotic cells is shifted, reflecting an average of seven sides. **d**, Cells on the periphery of a *string*-expressing clone (green) in the peripodial epithelium have fewer sides. Note the presence of a triangular cell, not observed under normal circumstances (yellow arrow; Dlg, red). **e**, The distribution of polygon types on the periphery of *string*-expressing clones (red; $n = 295$ cells in 24 clones) is shifted to reflect a predominance of pentagons, and deviates from controls (green, $n = 411$ wild-type peripodial cells).

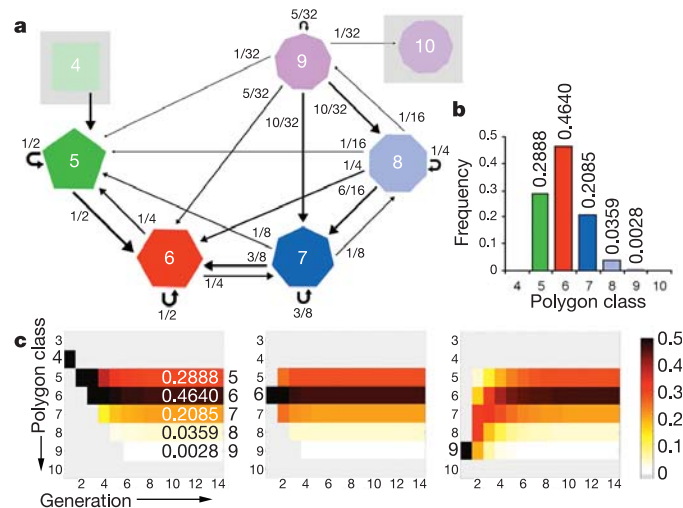


Figure 2 | A robust equilibrium topology in proliferating epithelial cell networks. **a**, Schematized Markov chain model representing proliferation dynamics for polygonal cells. Cells occupy a series of discrete states representing polygon classes from four to nine sides. To simulate a round of division, cells either recycle to the same state (curved arrows) or transition to a new state (straight arrows) according to the probabilities encoded by the neighbour effect transition matrix T (expressed here as fractions). Transitions from states s with an equilibrium probability $p_s < 10^{-4}$ have been omitted for clarity. **b**, Within ten generations, the distribution of polygonal shapes converges to E , comprising 28.88% pentagons, 46.40% hexagons, 20.85% heptagons, 3.59% octagons and 0.28% nonagons. **c**, Emergence of E regardless of initial conditions where all cells are uniformly quadrilateral, hexagonal or nonagonal.

One consequence of this result is that localized deviations in the rate of cell proliferation should predictably distort local topology. To test this, we took advantage of the fact that peripodial cells of the wing disc maintain the distribution *E*, but cease dividing by the mid-third larval instar. By misexpressing the mitosis-promoting phosphatase *string*²⁵ we drove peripodial cell clones through extra divisions, creating sharp boundaries between over-proliferating cells and their quiescent neighbours (Fig. 3d). Consistent with our model, dividing cells bounded by quiescent cells had fewer sides than controls (an average of 5.42 ± 0.14 sides compared with 5.94 ± 0.15 , respectively; Fig. 3e). In fact, three of 24 *string*-expressing peripodial clones contained triangular cells (Fig. 3d, yellow arrow), which were not observed in wild-type cells. These results demonstrate an unforeseen mechanism by which differential proliferation could influence cell shape and morphogenesis, or alternatively, cause the dysplastic tissue architecture observed in various forms of proliferative disease.

We propose that the division mechanism of adherent epithelial cells is not only mathematically sufficient to explain the predominantly hexagonal topology of epithelia, but also to predict the overall distribution of polygonal cell types. As a result, epithelial topology is irregular, but not random. Our results indicate that a simple emergent mechanism determines cell shape, suggesting a means by which epithelia accommodate rapid proliferation while maintaining uniform structural integrity. Looking forward, the Markov model formulation provides a new framework for investigating other models of cell division, such as different cleavage plane choices or aberrant cell division. This may be of general utility in understanding how stochastic behaviour at the single cell level manifests in global patterns in a multicellular context.

METHODS

Confocal time lapse. We used standard Ringer's solution (130 mM NaCl, 5 mM KCl, 1.5 mM MgCl₂) as well as Shields and Sang M3 insect media (Sigma-Aldrich; modified with 10% fetal bovine serum, 10 mU l⁻¹ insulin, 10 U ml⁻¹ penicillin, 10 µg ml⁻¹ streptomycin) to obtain movies over culture periods of 1.5–2 h at maximum.

Clonal analysis and imaging. GFP-expressing clones were induced in flies of the genotype: *yw hs-flp*¹²²; *Actin5c* >> *Gal4*, *UAS-GFP/+* with a 15-min heat shock at 37°C followed by a 10-h recovery period. *string*-expressing clones were induced in the same genotype with *UAS-sg* (Bloomington Stock Center) crossed onto chromosome III. For immunocytochemistry and phalloidin staining, *Drosophila* imaginal discs, whole *Hydra* and *Xenopus* tail epidermis were fixed in 4% paraformaldehyde in PBS. Images were collected on a Leica TCS SP2 AOBs confocal microscope system and processed using Adobe Photoshop 7.0 software.

Polygon distributions. Polygon distributions were determined by eye in confocal micrographs; error was estimated as the average standard deviation between counts from different images. Empirically, it was not possible to account systematically for certain rare but inevitable irregularities in real epithelia, such as occasional four-way point junctions and dying or grossly misshapen cells. The raw counts for cells of different sidedness are as follows: *Drosophila* disc columnar epithelium (4, 64; 5, 606; 6, 993; 7, 437; 8, 69; 9, 3). *Hydra* (4, 16; 5, 159; 6, 278; 7, 125; 8, 23; 9, 1). *Xenopus* (3, 2; 4, 40; 5, 305; 6, 451; 7, 191; 8, 52; 9, 8; 10, 2). *Drosophila* peripodial controls (4, 11; 5, 106; 6, 198; 7, 86; 8, 10; 9, 0). *Drosophila* peripodial *string* clones (only cells on the clone periphery were scored: 3, 3; 4, 27; 5, 134; 6, 105; 7, 21; 8, 5; 9, 0).

Markov chain convergence and equilibrium distribution calculation. The Perron–Frobenius theorem²⁴ guarantees that an irreducible and aperiodic Markov chain will converge to a unique stable equilibrium **p**^{*}, where **p**^{*} is the principal eigenvector of the transition matrix *T*. We truncated the infinite transition matrix *T* down to 20 rows and 20 columns and then used Matlab to calculate **p**^{*}. We also computed λ₂, the second-largest eigenvalue of *T*, as 0.5, which determines the rate of convergence.

Received 13 March; accepted 23 June 2006.

Published online 9 August 2006.

1. Schwann, T. *Microscopical Researches into the Accordance of Structure and Growth in Animals and Plants* (Syndenham Society, London, 1847).

2. Thompson, D. W. *On Growth and Form* (Cambridge Univ. Press, Cambridge, 1942).
3. Hayashi, T. & Carthew, R. W. Surface mechanics mediate pattern formation in the developing retina. *Nature* **431**, 647–652 (2004).
4. Amonlirdviman, K. et al. Mathematical modeling of planar cell polarity to understand domineering nonautonomy. *Science* **307**, 423–426 (2005).
5. Weaire, D. & Rivier, N. Soap, cells and statistics—Random patterns in two dimensions. *Contemp. Phys.* **25**, 59–99 (1984).
6. Abbott, L. A. & Lindenmayer, A. Models for growth of clones in hexagonal cell arrangements: Applications in *Drosophila* wing disc epithelia and plant epidermal tissues. *J. Theor. Biol.* **90**, 495–544 (1981).
7. Lewis, F. T. The effect of cell division on the shape and size of hexagonal cells. *Anat. Rec.* **33**, 331–355 (1926).
8. Lewis, F. T. The correlation between cell division and the shapes and sizes of prismatic cells in the epidermis of *Cucumis*. *Anat. Rec.* **38**, 341–376 (1928).
9. Graustein, W. C. On the average number of sides of polygons of a net. *Ann. Math.* **32**, 149–153 (1931).
10. Garcia-Bellido, A. & Merriam, J. R. Parameters of the wing imaginal disc development in *Drosophila melanogaster*. *Dev. Biol.* **24**, 61–87 (1971).
11. Bryant, P. J. & Levinson, P. Intrinsic growth control in the imaginal primordia of *Drosophila*, and the autonomous action of a lethal mutation causing overgrowth. *Dev. Biol.* **107**, 355–363 (1985).
12. Milan, M., Campuzano, S. & Garcia-Bellido, A. Cell cycling and patterned cell proliferation in the wing primordium of *Drosophila*. *Proc. Natl Acad. Sci. USA* **93**, 640–645 (1996).
13. Tepass, U., Tanentzapf, G., Ward, R. & Fehon, R. Epithelial cell polarity and cell junctions in *Drosophila*. *Annu. Rev. Genet.* **35**, 747–784 (2001).
14. Bertet, C., Sulak, L. & Lecuit, T. Myosin-dependent junction remodelling controls planar cell intercalation and axis elongation. *Nature* **429**, 667–671 (2004).
15. Zallen, J. A. & Zallen, R. Cell-pattern disordering during convergent extension in *Drosophila*. *J. Phys. Condens. Matter* **16**, S5073–S5080 (2004).
16. Classen, A. K., Anderson, K. I., Marois, E. & Eaton, S. Hexagonal packing of *Drosophila* wing epithelial cells by the planar cell polarity pathway. *Dev. Cell* **9**, 805–817 (2005).
17. Morin, X., Daneman, R., Zavortink, M. & Chia, W. A protein trap strategy to detect GFP-tagged proteins expressed from their endogenous loci in *Drosophila*. *Proc. Natl Acad. Sci. USA* **98**, 15050–15055 (2001).
18. Kello, R. J. et al. Flytrap, a database documenting a GFP protein-trap insertion screen in *Drosophila melanogaster*. *Nucleic Acids Res.* **32**, D418–D420 (2004).
19. Bryant, P. J. Cell lineage relationships in the imaginal wing disc of *Drosophila melanogaster*. *Dev. Biol.* **22**, 389–411 (1970).
20. Garcia-Bellido, A., Ripoll, P. & Morata, G. Developmental compartmentalisation of the wing disk of *Drosophila*. *Nat. New Biol.* **24**, 251–253 (1973).
21. Resino, J., Salama-Cohen, P. & Garcia-Bellido, A. Determining the role of patterned cell proliferation in the shape and size of the *Drosophila* wing. *Proc. Natl Acad. Sci. USA* **99**, 7502–7507 (2002).
22. Struhl, G. & Basler, K. Organizing activity of wingless protein in *Drosophila*. *Cell* **72**, 527–540 (1993).
23. Cowan, R. & Morris, V. B. Division rules for polygonal cells. *J. Theor. Biol.* **131**, 33–42 (1988).
24. Taylor, H. M. & Karlin, S. *An Introduction to Stochastic Modeling* 3rd edn (Academic, Chestnut Hill, Massachusetts, 1998).
25. Edgar, B. A. & O'Farrell, P. H. The three postblastoderm cell cycles of *Drosophila* embryogenesis are regulated in G2 by *string*. *Cell* **62**, 469–480 (1990).
26. Halanych, K. M. & Passamaneck, Y. A brief review of metazoan phylogeny and future prospects in Hox-research. *Am. Zool.* **41**, 629–639 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Ward, J. D. Lambert, B. Mathey-Prevot, E. Lieberman, R. Arnaout and M. Markstein for critical comments on the manuscript; the Bloomington Stock Center for fly stocks; and the Developmental Studies Hybridoma Bank for antibodies. This work was supported by a Microsoft New Faculty Fellowship to R.N., an NSF grant to R.N., and support from the Howard Hughes Medical Institute to N.P. A.B.P. is a fellow of the National Science Foundation and M.C.G. was supported by the Jane Coffin Childs Memorial Fund for Medical Research.

Author Contributions R.N. and N.P. are senior authors. A.B.P. and M.C.G. contributed equally to the work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.C.G. (mgx@stowers-institute.org) or R.N. (rad@eecs.harvard.edu).

LETTERS

Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans

Mathias Pessiglione¹, Ben Seymour¹, Guillaume Flandin¹, Raymond J. Dolan¹ & Chris D. Frith¹

Theories of instrumental learning are centred on understanding how success and failure are used to improve future decisions¹. These theories highlight a central role for reward prediction errors in updating the values associated with available actions². In animals, substantial evidence indicates that the neurotransmitter dopamine might have a key function in this type of learning, through its ability to modulate cortico-striatal synaptic efficacy³. However, no direct evidence links dopamine, striatal activity and behavioural choice in humans. Here we show that, during instrumental learning, the magnitude of reward prediction error expressed in the striatum is modulated by the administration of drugs enhancing (3,4-dihydroxy-L-phenylalanine; L-DOPA) or reducing (haloperidol) dopaminergic function. Accordingly, subjects treated with L-DOPA have a greater propensity to choose the most rewarding action relative to subjects treated with haloperidol. Furthermore, incorporating the magnitude of the prediction errors into a standard action-value learning algorithm accurately reproduced subjects' behavioural choices under the different drug conditions. We conclude that dopamine-dependent modulation of striatal activity can account for how the human brain uses reward prediction errors to improve future decisions.

Dopamine is closely associated with reward-seeking behaviours, such as approach, consummation and addiction^{3–5}. However, exactly how dopamine influences behavioural choice towards available rewards remains poorly understood. Substantial evidence from experiments on primates has led to the hypothesis that midbrain dopamine cells encode errors in reward prediction, the 'teaching signal' embodied in modern computational reinforcement learning theory⁶. Accumulating data indicate that different aspects of the dopamine signal incorporate information about the time, context, probability and magnitude of an expected reward^{7–9}. Furthermore, dopamine terminal projections are able to modulate the efficacy of cortico-striatal synapses^{10,11}, providing a mechanism for the adaptation of striatal activities during learning. Thus, dopamine-dependent plasticity could explain how striatal neurons learn to represent both upcoming reward and optimal behaviour^{12–16}. However, no direct evidence is available that links dopamine, striatal plasticity and reward-seeking behaviour in humans. More specifically, although striatal activity has been closely associated with instrumental learning in humans^{17,18}, there is no evidence that this activity is modulated by dopamine. Here we establish this link by using combined behavioural, pharmacological, computational and functional magnetic resonance imaging techniques.

We assessed the effects of haloperidol (an antagonist of dopamine receptors) and L-DOPA (a metabolic precursor of dopamine) on both brain activity and behavioural choice in groups of healthy subjects. Subjects performed an instrumental learning task involving monetary gains and losses, which required choosing between two novel visual stimuli displayed on a computer screen, so as to

maximize payoffs (Fig. 1a). Each stimulus was associated with a certain probability of gain or loss: one pair of stimuli was associated with gains (£1 or nothing), a second pair was associated with loss (−£1 or nothing), and a third pair was associated with no financial outcomes. Thus, the first pair was designed to assess the effects of the drugs on the ability to learn the most rewarding choice. The second pair was a control condition for the specificity of drug effects, because it required subjects to learn from punishments (losses) instead of rewards (gains), with the same relative financial interests. The third

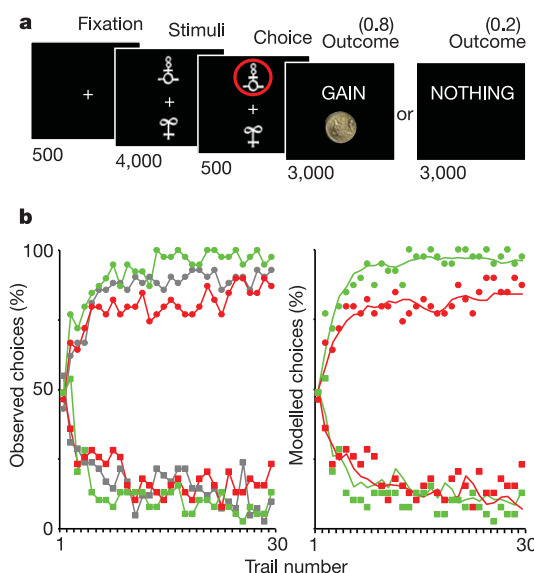


Figure 1 | Experimental task and behavioural results. **a**, Experimental task. Subjects selected either the upper or lower of two abstract visual stimuli presented on a display screen, and subsequently observed the outcome. In this example, the chosen stimulus is associated with a probability of 0.8 of winning £1 and a probability of 0.2 of winning nothing. Durations of the successive screens are given in milliseconds. **b**, Behavioural results. Left: observed behavioural choices for initial placebo (grey), superimposed over the results from the subsequent drug groups: L-DOPA (green) and haloperidol (red). The learning curves depict, trial by trial, the proportion of subjects that chose the 'correct' stimulus (associated with a probability of 0.8 of winning £1) in the gain condition (circles, upper graph), and the 'incorrect' stimulus (associated with a probability of 0.8 of losing £1) in the loss condition (squares, lower graph). Right: modelled behavioural choices for L-DOPA (green) and haloperidol (red) groups. The learning curves represent the probabilities predicted by the computational model. Circles and squares representing observed choices have been left for the purpose of comparison. All parameters of the model were the same for the different drug conditions, except the reinforcement magnitude R , which was estimated from striatal BOLD response.

¹Wellcome Department of Imaging Neuroscience, 12 Queen Square, London WC1N 3BG, UK.

pair was a neutral condition allowing further control, in which subjects could indifferently choose any of the two stimuli, because they involved no monetary gain or loss. The probabilities were reciprocally 0.8 and 0.2 in all three pairs of stimuli, which were randomly displayed in different trials within the same learning session. Subjects had to press a button to select the upper stimulus, or do nothing to select the lower stimulus, as they appeared on the display screen. This Go/NoGo mode of response offers the possibility of identifying brain areas related to motor execution of the choice, by contrasting Go and NoGo trials.

We first investigated the performance of a placebo-treated group, which showed that subjects learn within a 30-trial session to select the high-probability gain and avoid the high-probability loss. The overall performance was similar across the gain and loss trials, but with significantly lower inter-trial consistency and longer response time for the loss condition (Supplementary Table 1). This result indicates the possible existence of physiological differences between selecting actions to achieve rewards and selecting actions to avoid losses, possibly corresponding to additional processes being recruited during the avoidance condition. In the subsequent pharmacological study, L-DOPA-treated subjects won more money than haloperidol-treated subjects (£66.7 $\pm$ 1.00 versus £61.0 $\pm$ 2.10 (errors indicate s.e.m.), $P < 0.05$), but did not lose less money (£26.7 $\pm$ 1.50 versus £28.9 $\pm$ 1.40). Thus, relative to haloperidol, L-DOPA increased the frequency which subjects chose high-probability gain but not the frequency which they chose low-probability loss (Fig. 1b). In other words, enhancing central dopaminergic activity improved choice performance towards monetary gains but not avoidance of monetary losses. Neither drug significantly influenced response times, percentages of Go responses or subjective ratings of mood, feelings and sensations (Supplementary Table 2 and Supplementary Fig. 1).

For the analysis of brain activity, we first examined the representation of outcome prediction errors across all groups (placebo, L-DOPA and haloperidol). Corresponding brain regions were identified in a linear regression analysis, conducted across all trials, sessions and subjects, with the prediction errors generated from a standard action-value learning model. The parameters were adjusted to maximize the likelihood of the subjects' choices under the model. For each trial the model calculated choice probabilities according to action values. After each trial the value of the chosen action was updated in proportion to the prediction error, defined as the difference between expected value and actual outcome.

Statistical parametric maps (SPMs) revealed large clusters that were positively correlated with reward prediction error, all located in the striatum: predominantly the bilateral ventral striatum and left posterior putamen (Fig. 2a). This appetitive prediction error was observed across both gain and loss conditions, indicating that the striatum might represent successfully avoided outcomes as relative rewards. In addition, we observed a cluster showing significant negative correlation with an appetitive prediction error during the loss (but not gain) trials in the right anterior insula. This corresponds to an aversive prediction error, indicating that the loss condition might engage opponent appetitive and aversive processes, an idea in keeping with an experimental psychological literature on the dual excitatory and inhibitory mechanisms involved in signalled avoidance learning¹⁹.

To characterize further the brain activity involved in behavioural choice, we next examined the main contrasts between trial types at the time of stimuli display (Fig. 2b). Bilateral ventral striatum was significantly activated in the contrast between gain and neutral stimuli, and also in the contrast between loss and neutral stimuli. This activity is consistent with a learned value reflecting the distinction between stimuli predicting gains or losses on the one hand, and those predicting mere neutral outcomes on the other. Again, the similarity of the signal across both gain and loss trials might indicate a comparable appetitive representation of stimuli predicting reward and punishment avoidance. The left posterior putamen was significantly

activated when the optimal stimulus was on the top of the screen rather than the bottom. This indicates that this region might be involved specifically when the optimal choice requires a Go (button press) and not a NoGo response. The left lateralization of posterior putamen activity is consistent with the fact that the right hand was employed for pressing the button. These findings are in line with a body of literature implicating the anterior ventral striatum in reward prediction^{20,21} and the posterior putamen in movement execution^{22,23}. The distinct functional roles that we ascribe to these striatal regions are also supported by their principal afferents^{24,25}: amygdala, orbital and medial prefrontal cortex for the ventral striatum versus somatosensory, motor and premotor cortex for the posterior putamen. The bilateral anterior insula was activated in the contrast between loss and neutral pairs alone, again providing support for the existence of an opponent aversive representation of stimulus value during avoidance learning. This same region of anterior insula has been shown to encode aversive cue-related prediction errors during pavlovian learning of physical punishment²⁶.

Last, we explored the effects of drugs (L-DOPA and haloperidol) on the representation of outcome prediction errors. We averaged the blood-oxygen-level-dependent (BOLD) responses over clusters reflecting prediction errors (derived from the above analysis), separately for the different drugs and outcomes (Fig. 3). Note that in striatal

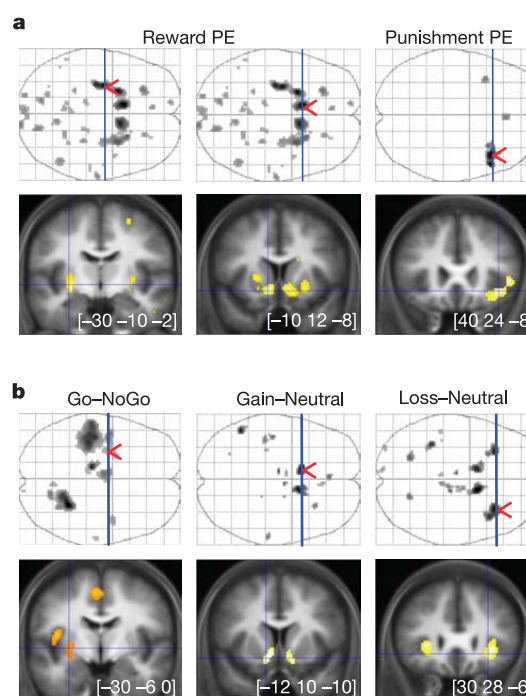


Figure 2 | Statistical parametric maps of prediction error and stimulus-related activity. Coronal slices (bottom) were taken at local maxima of interest indicated by red arrows on the axial projection planes (top). Areas shown in grey on axial planes and in orange or yellow on coronal slices showed significant effect after family-wise error correction for multiple comparisons ($P < 0.05$). **a**, Brain activity correlated with prediction errors derived from the computational model. Reward prediction errors (positive correlation) were found by conjunction of gain and loss conditions (left panels), whereas punishment prediction errors (negative correlation) were found in the loss condition alone (right panel). From left to right, MNI (Montreal Neurological Institute) coordinates are given for the maxima found in the left posterior putamen, left ventral striatum and right anterior insula. **b**, Statistical parametric maps resulting from main contrasts between stimuli conditions. Go and NoGo refer to stimuli position requiring, or not requiring, a button press to get the optimal outcome. Gain, neutral and loss correspond to the different pairs of stimuli. As above, the maxima shown are located in the left posterior putamen, left ventral striatum and right anterior insula, from left to right.

clusters the average amplitude of the negative BOLD response was about fourfold that of positive BOLD response, which was consistent with the expression of appetitive prediction error (converging towards +0.2 and -0.8). The right anterior insula showed the opposite pattern (during the loss trials), which was consistent with the expression of an aversive prediction error (converging towards +0.8 and -0.2). Comparing between drugs, there was a significant difference ($P < 0.05$) in the gain condition alone, with positive and negative BOLD responses being enhanced under L-DOPA in comparison with haloperidol. There was no significant effect in the loss condition, either in the striatum or in the anterior insula, in accord with the absence of drug effects on behavioural choices.

The asymmetry of drug effects between gain and loss conditions supports the hypothesis that striatal dopamine has a specific involvement in reward learning, providing new insight into the debate over its relative reward selectivity²⁷ given evidence implicating dopamine involvement in salient and aversive behaviours²⁸. In some paradigms, such as in the aversive component of various cognitive procedural learning tasks, dopamine depletion improves performance¹³. In others, however, such as conditioned avoidance response learning, dopamine blockade impairs performance²⁹, probably as a result of interference with appetitive processes underlying the opponent 'safety state' of the avoided outcome. Although our data support the expression of distinct appetitive and aversive prediction errors during avoidance learning, the fact that neither of these opponent signals was affected by the dopamine-modulating drugs leaves it still unclear precisely what function dopamine has in aversive instrumental learning. This uncertainty is confounded to some extent by the fact that we do not know unequivocally how the drugs affect the different components of dopaminergic function, for example with

regard to tonic versus phasic firing, or D1 versus D2 receptors. Thus, although we can assert that dopamine has a selective effect on gain-related striatal prediction errors, we have to be cautious about inferring the precise mechanism at a cellular level.

We then investigated whether there was any relationship between dopamine-modulated striatal activity and behaviour, during the gain condition. We first estimated the effective monetary reward value from the amplitude of the striatal BOLD responses, for the drug conditions in comparison with the placebo group. By taking the difference between positive and negative BOLD responses as equivalent to £1.00 for the placebo group, we estimated an effective reward value of £1.29 $\pm$ 0.07 under L-DOPA and £0.71 $\pm$ 0.12 under haloperidol. These values were within the 95% confidence interval of those provided by the maximum-likelihood estimate of the observed choices under our computational model (see Supplementary Fig. 2). In other words, when we incorporated the reward magnitudes estimated from striatal BOLD responses into the computational model, it accurately and specifically reproduced the effects of the drugs on behavioural choices (Fig. 1b).

Our results support a key functional link between dopamine, striatal activity and reward-seeking behaviour in humans. We have shown first that dopamine-related drugs modulate reward prediction errors expressed in the striatum, and second that the magnitude of this modulation is sufficient for a standard action-value learning model to explain the effects of drugs on behavioural choices. These findings suggest that humans use dopamine-dependent prediction errors to guide their decisions, and, more specifically, that dopamine modulates the apparent value of rewards as represented in the striatum. Furthermore, the findings might provide insight into models of clinical disorders in which dopamine is implicated, and for which L-DOPA and haloperidol are used as therapeutic agents, such as Parkinson's disease and schizophrenia. For example, it offers a potential mechanism for the development of compulsive behaviours (such as overeating, hypersexuality and pathological gambling) induced by dopamine replacement therapy in patients with Parkinson's disease³⁰.

METHODS

For a detailed and referenced description of the experimental and analytical techniques, see Supplementary Methods and Results.

Experimental procedure. The study was approved by the local ethics committee. In all, 39 healthy subjects were scanned (19–37 years old; 23 males), including a single-blind initial study of 13 subjects treated with a placebo only (lactose) and a double-blind test study of 26 subjects, half treated with Haldol (haloperidol, 1 mg) and half with Madopar (L-DOPA, 100 mg, plus benserazide, 25 mg). After a short practice, subjects had to perform three sessions of the same instrumental learning task, each proposing three new pairs of abstract stimuli. Each of the pairs of stimuli (gain, loss and neutral) was associated with pairs of outcomes ('gain' £1/nil, 'loss' £1/nil, 'look' £1/nil), the two stimuli corresponding to reciprocal probabilities (0.8/0.2 and 0.2/0.8). On each trial, one pair was randomly presented and the two stimuli were displayed on the screen, above and below a central fixation cross, their relative position being counterbalanced across trials. The subject was required to choose the upper stimulus by pressing a button (Go response), or the lower stimulus by doing nothing (NoGo response). The choice was then circled in red and the outcome was displayed on the screen. To win money the subjects had to learn, by trial and error, the stimulus–outcome associations. They were told that their winnings would be their remuneration for participation, but they all left with the same fixed amount. To assess for side effects of the drug, they were finally asked to rate their subjective feelings, using visual analogue scales. Behavioural performance was compared directly between the L-DOPA and haloperidol groups, with two-sample *t*-tests.

Computational model. A standard algorithm of action-value learning was then fitted to the observed behaviour. For each pair of stimuli A and B, the model estimates the expected values of choosing A (Q_A) and choosing B (Q_B), on the basis of individual sequences of choices and outcomes. The expected values were set at zero before learning, and after every trial $t > 0$ the value of the chosen stimulus (say A) was updated according to the rule $Q_A(t+1) = Q_A(t) + \alpha \delta(t)$. The outcome prediction error, $\delta(t)$, is the difference between the actual and the expected outcome, $\delta(t) = R(t) - Q_A(t)$, the reinforcement $R(t)$ being either +£1, £0 or -£1. Given the expected values, the probability (or likelihood) of the

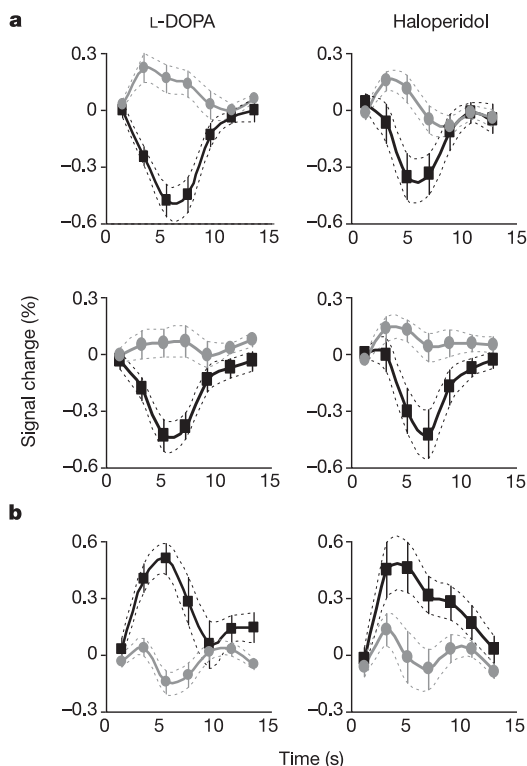


Figure 3 | Time course of brain responses reflecting prediction errors. Time courses were averaged across trials throughout the entire learning sessions. Error bars are inter-subject s.e.m. **a**, Overlaid positive (grey circles) and negative (black squares) reward prediction errors in the striatum for both L-DOPA-treated and haloperidol-treated groups, and in both gain and loss trials. **b**, Overlaid positive (black squares) and negative (grey circles) punishment prediction errors in the right anterior insula, during the loss trials.

observed choice was estimated with the softmax rule: $P_a(t) = \exp(Q_a(t)/\beta) / \{\exp[Q_a(t)/\beta] + \exp[Q_b(t)/\beta]\}$. The constants α (learning rate) and β (temperature) were adjusted to maximize the likelihood of the actual choices under the model, across all groups of subjects. Outcome prediction errors estimated by the model were then used as a statistical regressor in the imaging data.

Image acquisition and analysis. T_2^* -weighted echo planar images (EPIs) were acquired with BOLD contrast on a 3.0-T Siemens Allegra magnetic resonance scanner, using a tilted plane acquisition sequence covering the whole brain. T_1 -weighted structural images were normalized and averaged across subjects to allow group-level anatomical localization. EPIs were analysed in an event-related manner, with the statistical parametric mapping software SPM5. Preprocessing consisted of spatial realignment, normalization to a standard EPI template, and spatial smoothing with a 6-mm gaussian kernel. To correct for motion artefacts, subject-specific realignment parameters were modelled as covariates of no interest. Onsets of stimuli and outcomes were modelled as separate delta functions and convolved with a canonical haemodynamic response function. Prediction errors generated by the computational model were used as parametric modulation of additional regressors modelled at outcome onsets. Linear contrasts of regression coefficients were computed at the individual subject level and then taken to group-level t -tests. All group-level SPMs are reported with a threshold of $P < 0.05$ after family-wise error correction for the entire brain. For large clusters (more than 64 voxels) showing statistical covariation with the theoretical prediction error, the response time courses were estimated, with the use of a flexible basis set of finite impulse responses (FIRs), separated from the next by one scan (1.95 s). The area between positive and negative FIRs, over 3–9 s after outcome, were used to estimate effective reward values under the drug conditions.

Received 8 June; accepted 7 July 2006.

Published online 23 August 2006.

- Thorndike, E. L. *Animal Intelligence: Experimental Studies* (Macmillan, New York, 1911).
- Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
- Wise, R. A. Dopamine, learning and motivation. *Nature Rev. Neurosci.* **5**, 483–494 (2004).
- Everitt, B. J. *et al.* Associative processes in addiction and reward. The role of amygdala-ventral striatal subsystems. *Ann. NY Acad. Sci.* **877**, 412–438 (1999).
- Ikemoto, S. & Panksepp, J. The role of nucleus accumbens dopamine in motivated behavior: a unifying interpretation with special reference to reward-seeking. *Brain Res. Brain Res. Rev.* **31**, 6–41 (1999).
- Hollerman, J. R. & Schultz, W. Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature Neurosci.* **1**, 304–309 (1998).
- Waelti, P., Dickinson, A. & Schultz, W. Dopamine responses comply with basic assumptions of formal learning theory. *Nature* **412**, 43–48 (2001).
- Nakahara, H., Itoh, H., Kawagoe, R., Takikawa, Y. & Hikosaka, O. Dopamine neurons can represent context-dependent prediction error. *Neuron* **41**, 269–280 (2004).
- Bayer, H. M. & Glimcher, P. W. Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron* **47**, 129–141 (2005).
- Smith, A. D. & Bolam, J. P. The neural network of the basal ganglia as revealed by the study of synaptic connections of identified neurones. *Trends Neurosci.* **13**, 259–265 (1990).
- Calabresi, P. *et al.* Synaptic transmission in the striatum: from plasticity to neurodegeneration. *Prog. Neurobiol.* **61**, 231–265 (2000).
- Tremblay, L., Hollerman, J. R. & Schultz, W. Modifications of reward expectation-related neuronal activity during learning in primate striatum. *J. Neurophysiol.* **80**, 964–977 (1998).
- Frank, M. J., Seeberger, L. C. & O'Reilly, R. C. By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science* **306**, 1940–1943 (2004).
- Hollerman, J. R., Tremblay, L. & Schultz, W. Influence of reward expectation on behavior-related neuronal activity in primate striatum. *J. Neurophysiol.* **80**, 947–963 (1998).
- Lauwereyns, J., Watanabe, K., Coe, B. & Hikosaka, O. A neural correlate of response bias in monkey caudate nucleus. *Nature* **418**, 413–417 (2002).
- Samejima, K., Ueda, Y., Doya, K. & Kimura, M. Representation of action-specific reward values in the striatum. *Science* **310**, 1337–1340 (2005).
- O'Doherty, J. *et al.* Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* **304**, 452–454 (2004).
- Tanaka, S. C. *et al.* Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nature Neurosci.* **7**, 887–893 (2004).
- Dickinson, A. *Contemporary Animal Learning Theory* (Cambridge Univ. Press, Cambridge, 1980).
- O'Doherty, J. P., Deichmann, R., Critchley, H. D. & Dolan, R. J. Neural responses during anticipation of a primary taste reward. *Neuron* **33**, 815–826 (2002).
- Knutson, B., Taylor, J., Kaufman, M., Peterson, R. & Glover, G. Distributed neural representation of expected value. *J. Neurosci.* **25**, 4806–4812 (2005).
- Jueptner, M. & Weiller, C. A review of differences between basal ganglia and cerebellar control of movements as revealed by functional imaging studies. *Brain* **121**, 1437–1449 (1998).
- Lehericy, S. *et al.* Motor control in basal ganglia circuits using fMRI and brain atlas approaches. *Cereb. Cortex* **16**, 149–161 (2006).
- Alexander, G. E., DeLong, M. R. & Strick, P. L. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu. Rev. Neurosci.* **9**, 357–381 (1986).
- Haber, S. N. The primate basal ganglia: parallel and integrative networks. *J. Chem. Neuroanat.* **26**, 317–330 (2003).
- Seymour, B. *et al.* Temporal difference models describe higher-order learning in humans. *Nature* **429**, 664–667 (2004).
- Ungless, M. A., Magill, P. J. & Bolam, J. P. Uniform inhibition of dopamine neurons in the ventral tegmental area by aversive stimuli. *Science* **303**, 2040–2042 (2004).
- Salamone, J. D. The involvement of nucleus accumbens dopamine in appetitive and aversive motivation. *Behav. Brain Res.* **61**, 117–133 (1994).
- Cook, L., Morris, R. W. & Mattis, P. A. Neuropharmacological and behavioral effects of chlorpromazine (thorazine hydrochloride). *J. Pharmacol. Exp. Ther.* **113**, 11–12 (1955).
- Molina, J. A. *et al.* Pathologic gambling in Parkinson's disease: a behavioral manifestation of pharmacologic treatment? *Mov. Disord.* **15**, 869–872 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Friston for discussions, B. Draganski for assistance in the double-blind procedure, and J. Daunizeau for assistance in the statistical analysis. This work was funded by the Wellcome Trust research programme grants. M.P. received a grant from the Fyssen Foundation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.P. (m.pessiglione@filion.ucl.ac.uk).

LETTERS

Transgeneration memory of stress in plants

Jean Molinier¹†, Gerhard Ries¹†, Cyril Zipfel¹† & Barbara Hohn¹

Owing to their sessile nature, plants are constantly exposed to a multitude of environmental stresses to which they react with a battery of responses. The result is plant tolerance to conditions such as excessive or inadequate light, water, salt and temperature, and resistance to pathogens. Not only is plant physiology known to change under abiotic or biotic stress, but changes in the genome have also been identified^{1–5}. However, it was not determined whether plants from successive generations of the original, stressed plants inherited the capacity for genomic change. Here we show that in *Arabidopsis thaliana* plants treated with short-wavelength radiation (ultraviolet-C) or flagellin (an elicitor of plant defences⁶), somatic homologous recombination of a transgenic reporter is increased in the treated population and these increased levels of homologous recombination persist in the subsequent, untreated generations. The epigenetic trait of enhanced homologous recombination could be transmitted through both the maternal and the paternal crossing partner, and proved to be dominant. The increase of the hyper-recombination state in generations subsequent to the treated generation was independent of the presence of the transgenic allele (the recombination substrate under consideration) in the treated plant. We conclude that environmental factors lead to increased genomic flexibility even in successive, untreated generations, and may increase the potential for adaptation.

Plants are influenced by abiotic and biotic environmental factors on several levels; apart from changes in plant physiology and the mounting of resistance responses, the dynamics of the genome can also be altered. Examples include the activation of transposable elements by abiotic and biotic stress conditions^{7–9}, induction of mutations by chemical and physical agents¹⁰, and enhancement of homologous recombination by elevated temperatures¹¹ or ultraviolet-B (UV-B) (ref. 2). Especially interesting is the genomic flexibility shown by plant genomes in response to pathogen attack^{3,4,7}. Whenever possible, such changes were monitored at the level of the sequence of affected genes. The influence these changes have in evolutionary terms, however, remained poorly understood, because most changes were detected in somatic tissue and not considered in further generations. In plants, the reproductive cell-lineage emerges from somatic tissue late in development¹², thus some genomic changes acquired during the life of a plant can be transmitted to the next generation. Indeed, with progeny of UV-B- or pathogen-treated plants, the frequency of occurrence of genetically fixed mutation (in this case, homologous recombination) was reproducibly elevated^{2,4}. The degree of genomic change in the offspring of the stressed population was expected to return to the basal level. We show here that increased levels of homologous recombination persist for several generations in the lineage from the original parent plants that were exposed to stresses, including ultraviolet radiation or flagellin.

We measured the rate of homologous recombination in the untreated offspring of plants exposed to conditions of environmental

stress. We used *A. thaliana* plants harbouring β -glucuronidase (*GUS*)-based constructs in which truncated but overlapping parts of the gene allow quantification of somatic homologous recombination. The results of this event are visualized as blue spots on a white background following histochemical staining of plants (Fig. 1a, b). Previous molecular analyses of the plant DNA confirmed that the blue spots, which represent *GUS* activity, indeed symbolize bona fide recombination events^{13,14}. Using this assay, the influence of ultraviolet-C (UV-C) was tested in six independent transgenic lines that carried the recombination reporter in different relative orientations of the *GUS* sequence fragments: 'GU' and 'US' (ref. 15). The basal levels of homologous recombination, indicated as numbers of recombination sectors per plant, varied among the six lines; the degrees of stimulation were also different, but in all cases the treatment with UV-C stimulated the level of homologous recombination (Fig. 1c). UV-C induction of homologous recombination together with variation between independent transgenic lines is consistent with previous reports².

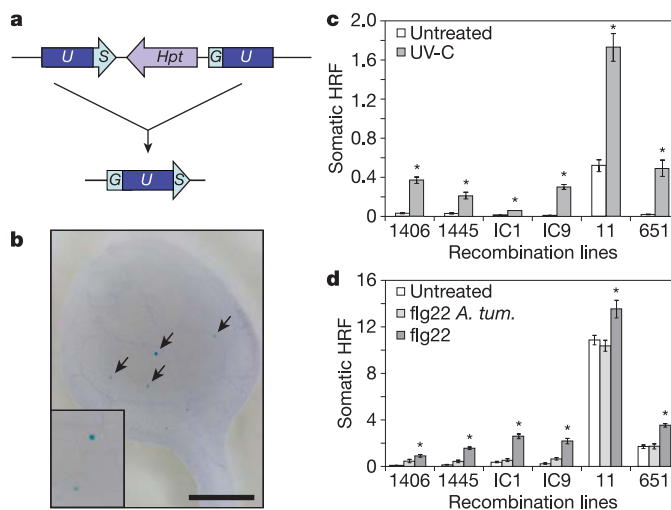


Figure 1 | Somatic homologous recombination in UV-C- and flg22-treated plants. **a**, Schematic representation of a recombination substrate used for monitoring somatic homologous recombination (lines IC1 and IC9). *GUS*, β -glucuronidase gene; *Hpt*, hygromycin-resistance gene. Homologous region is shown in dark blue. **b**, Recombination events (blue spots highlighted by black arrows) giving a measure of homologous recombination frequency (HRF; see Methods) in line IC1 after flg22 treatment. Scale bar, 1 mm; inset, X3 original magnification. **c**, Somatic HRF in untreated and UV-C-treated *S₀* plants. Results are means $\pm$ s.e.m. ($n > 50$ plants; t -test $*P < 0.05$). **d**, Somatic HRF in either untreated plants, plants treated with flg22 *A. tum.*, or treated with flg22. Results are means $\pm$ s.e.m. ($n > 40$ plants; t -test $*P < 0.05$).

¹Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, CH-4058 Basel, Switzerland. †Present addresses: Institut de Biologie Moléculaire des Plantes, 12 Rue du Général Zimmer, F-67084 Strasbourg Cedex, France (J.M.); BioMedinvestor AG, Elisabethenstrasse 23, CH-4051 Basel, Switzerland (G.R.); The Sainsbury Laboratory, John Innes Centre, Colney Lane, Norwich NR4 7UH, UK (C.Z.).

Inoculation of homologous-recombination-tester plants with pathogens was previously shown to increase homologous recombination frequencies^{3,4}. Here we use flagellin, a bacterium-derived elicitor, to activate plant defences. In *A. thaliana*, perception of flagellin, the building block of the bacterial flagellum, occurs by recognition of the most conserved domain on the flagellin amino terminus, represented by the peptide flg22 (ref. 16). Treatment of plants with this peptide has been shown to trigger resistance to pathogenic bacteria⁶, unlike treatment with the inactive analogue from *Agrobacterium tumefaciens* (flg22 *A. tum.*). In all six transgenic lines analysed, we detected an increase of homologous recombination on treatment with flg22, but not with the inactive flg22 *A. tum.* peptide (Fig. 1d). Thus, flagellin represents an agent mimicking a pathogen in inducing both a pathogenic response⁶ and elevated levels of homologous recombination.

Progeny of UV-treated and self-pollinated plants grown under non-stress conditions were analysed for homologous recombination. In all six lines, the frequency of homologous recombination was greater than that of plants from untreated progenitors (Fig. 2a). As the location of the transgene in the different lines varies, the transgenerational effect of increased frequencies of somatic homologous recombination is most likely independent of genome position. Homologous recombination frequencies in the S_1 selfed generation of treated plants were 2–4-fold higher than those of S_1 plants derived from untreated parents. The effect was also independent of ecotype and orientation

of the recombination substrate (see Methods for details), as lines 1406, 1445, IC1 and IC9 were in the Columbia ecotype, whereas lines 11 and 651 were in C24. Similarly, the progeny of flg22- and flg22 *A. tum.*-treated plants, and of untreated plants, were analysed for persistence of elevated levels of homologous recombination; again, the plants 'memorized' their previous exposure or reaction to the biologically active peptide flg22 and exhibited a constitutively elevated level of somatic recombination (Fig. 2b). This response was specific, as treatment with the inactive peptide did not lead to the described transgenerational effect.

The basis for the described transgenerational effect must be epigenetic because the whole population changes its behaviour, whereas a mutation would affect only very few plants. Epigenetic change can be described as mitotically and/or meiotically heritable, potentially reversible chromatin alteration, which occurs in the absence of change in the DNA sequence. Heritable changes of epigenetic traits may last for several generations. We tested whether the homologous-recombination-enhancing epigenetic change acquired through the environmental stimulus of ultraviolet radiation can be detected in generations succeeding S_1 . Somatic homologous recombination was persistently found to be increased up to the selfed S_4 progeny of plants that were treated with UV-C only in S_0 (Fig. 3). The epigenetic change leading to enhanced homologous recombination frequencies is thus stable for at least four generations.

So far, data on homologous recombination in successive generations following treatment were generated by analysing self-pollinated plants. It was of interest to test whether the epigenetic changes could be preferentially transmitted through the male or the female germline. To test this, plants were irradiated with UV-C and crossed to untreated plants. In offspring of treated and self-pollinated plants of line 1406, the frequency of homologous recombination was about one blue spot per plant as opposed to untreated plants, which gave rise to about 0.4 spots per plant in the next generation (Fig. 4a). However, recombination in the progeny of plants from a cross where only one parent was UV-treated was as high as in the treated, self-pollinated plants, irrespective of the direction of the cross. Using line 1445, a similar picture emerged (Fig. 4a). Comparable results were obtained in an experiment in which only one crossing partner was treated with flagellin; again, recombination was elevated to a similar level in the offspring of UV-treated outcrossed and selfed plants (Fig. 4b). We conclude that the epigenetic 'memory' could be inherited through both gametes in a dominant manner.

The epigenetic change revealed may be inscribed on the entire genome, on a particular locus, or on the transgene of the treated

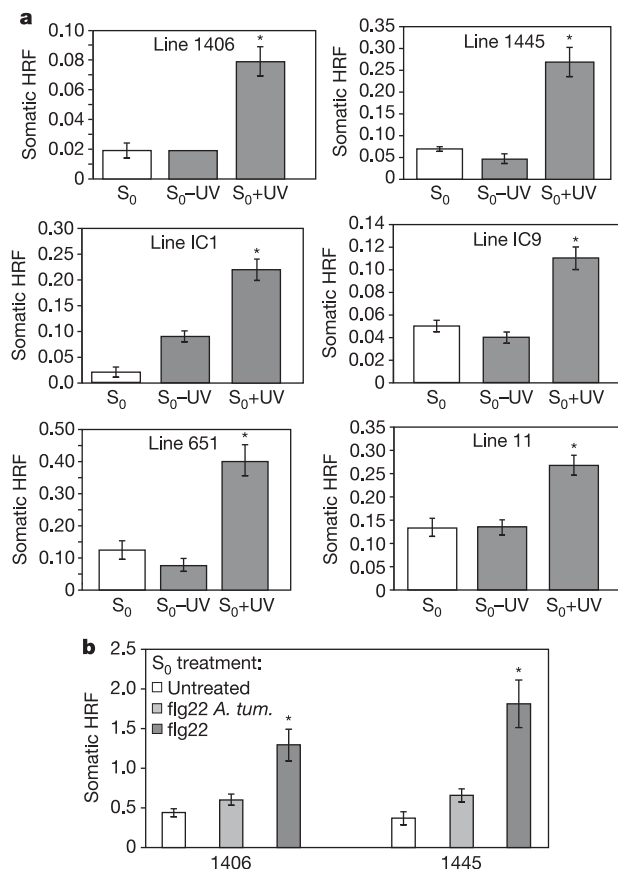


Figure 2 | Somatic HRF in S_0 lines and their respective S_1 progenies. **a**, Somatic HRF in six different untreated S_0 lines (white bars) and in untreated S_1 plants (grey bars) derived from those same S_0 lines treated with UV-C (S_0 + UV) or untreated (S_0 - UV). Results are means $\pm$ s.e.m. ($n > 40$ plants; t -test $*P < 0.05$ compared with S_1 plants from S_0 - UV). **b**, Somatic HRF in S_1 plants (lines 1406 and 1445) from S_0 plants that were untreated, treated with the inactive flagellin analogue (flg22 *A. tum.*), or treated with the active flagellin (flg22). Results are means $\pm$ s.e.m. ($n > 40$ plants; t -test $*P < 0.05$ compared with S_1 treated with flg22 *A. tum.*).

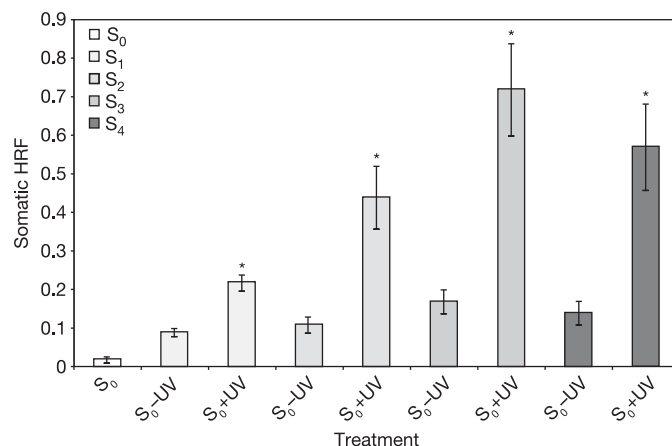


Figure 3 | Somatic HRF in S_0 plants and in the next four generations. S_0 plants (line IC1) were either untreated or UV-treated. Somatic HRF was measured in untreated S_1 , S_2 , S_3 and S_4 plants. Results are means $\pm$ s.e.m. ($n > 50$ plants; t -test $*P < 0.05$ compared with the corresponding S_0 - UV generation).

plants. Although the first two alternatives are difficult to address, we tested whether the irradiated plants have to contain the transgene in order to transmit their epigenetic status to the descending generation. We therefore crossed irradiated or untreated *A. thaliana* wild-type plants lacking the transgene locus with treated or untreated transgenic plants and analysed the somatic homologous recombination frequency of the offspring. When both parental plants were exposed to UV-C, recombination was elevated by a factor of about four in comparison with untreated parents (Fig. 4c, columns 1 and 2, 5 and 6). When an irradiated female transgenic plant was crossed to a non-irradiated male wild-type plant, homologous recombination was increased to a similar extent (column 3). Therefore, the irradiated plant does not require the presence of the transgene in the crossing partner in order to be dominant (this also confirms and extends the data shown above, in which only one parent was irradiated, but both were transgenic; Fig. 4a). Column 4 in Fig. 4c shows the reciprocal experiment: homologous recombination was measured in the offspring of plants in which only the non-transgenic crossing partner was exposed to irradiation, whereas the other partner carried the transgene. Comparison to columns 1, 2 and 3 allows the conclusion that the epigenetic change, measured as increased potential to undergo homologous recombination, can be transmitted through crossing with an irradiated plant lacking the recombination-substrate locus. Therefore, irradiation does not directly change the epigenetic status of the transgene locus, and activation (whatever this 'activation' may be) of the recombination potential of the transgene locus can be accomplished in *trans*. Again, the outcome of this experiment is independent of the direction of performed crosses: columns 5–8 show levels of homologous recombination measured in offspring of crosses in which the female parent was non-transgenic and the pollen-donor carried the transgene.

Influences of the environment on the plant genome have been documented, and include the activation of transposition in maize^{7,8}

and Solanaceae⁹; genomic rearrangement following changes in the nutrition of flax⁵ and climatic conditions in *Tradescantia*¹⁷; and alterations in the frequency of homologous recombination following exposure of plants to a variety of agents^{2–4,11,18}. These influences have been interpreted as 'genomic shock' (refs 1, 19, 20). Here we demonstrated that environmental influences, specifically ultraviolet radiation and a bacterial elicitor, change the flexibility of the plant genome in somatic tissue of treated plants and in somatic tissue of their progeny. As these influences persist in the entire population of plants, the basis for the change is epigenetic rather than genetic. Plants carrying the transgene locus do not have to face the environmental challenges themselves in order to transmit the epigenetic change to the offspring; the stimulus for an increase of recombination can be imposed in *trans* by a single treated parent. It is unclear, however, whether this stimulation is exerted by activation of a component of the homologous recombination machinery or by rendering the recombination locus more accessible, for instance by chromatin restructuring. Transcriptome analyses of *S*₀ and *S*₁ plants originating from untreated or treated (UV-C or flg22) plants did not reveal significant changes in gene expression (data not shown). This indicates that a global transcriptional stimulation of key players of the homologous recombination machinery may not be responsible for the observed phenomenon. However, it cannot be excluded that this functional genomic approach failed to reveal subtle transcriptional changes or induction of microRNAs responsible for the observed transgenerational 'memory' effect. A mechanism resembling paramutation in which one allele changes the epigenetic state of the other can be excluded; the enhancement of recombination can be exerted by a stressed crossing partner that lacks the transgene. The demonstrated epigenetic change is heritable for at least four generations. In this respect, it resembles epimutation such as that in *Linaria vulgaris* affecting floral symmetry²¹, or that in *A. thaliana*, called *SUPERMAN*, also affecting flower development²². Because of

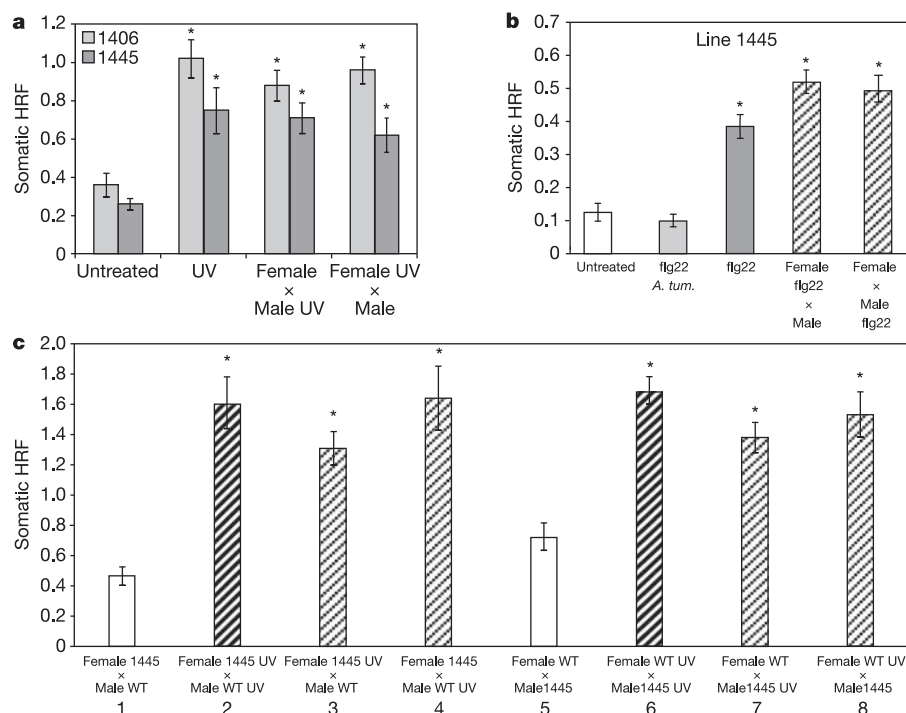


Figure 4 | Somatic HRF in either self-pollinated or outcrossed plants.

a, Somatic HRF in offspring of either self-pollinated untreated plants, UV-treated plants, or plants in which one of the parents was UV-treated ($n > 40$ plants; t -test $*P < 0.05$). **b**, Somatic HRF in offspring of either self-pollinated untreated plants (white bar), flg22-treated plants (grey bars), or plants in which one of the parents was flg22-treated (hatched bars; $n > 35$

plants; t -test $*P < 0.05$). **c**, Somatic HRF in offspring of plants in which one parent was wild type (WT) and the other harboured the recombination substrate ($n > 40$ plants; t -test $*P < 0.05$). White bar, both parents untreated; dark hatching, both parents UV-treated; light hatching, one parent UV-treated. Refer to the main text for the lines used. All the results are means $\pm$ s.e.m.

the genetic stability of the epigenetic change in recombination potential, it is conceivable that meiotic recombination is affected as well.

We have demonstrated an induced, epigenetic, heritable change of a molecularly defined and quantitatively measured trait. However, an adaptive value of the observed changes cannot easily be evaluated at this point. We propose that the environmental influences that lead to increased genomic dynamics even in successive, untreated generations may increase the potential for adaptive evolution²³. This work should help in elucidating the underlying molecular mechanism that causes the observed transgenerational 'memory'.

METHODS

Plant material. Homozygous *A. thaliana* lines, IC1 and IC9 (ref. 14), 1406 and 1445 (ref. 24) (ecotype Columbia), 11 and 651 (ref. 25) (ecotype C24), each carrying a single copy of a recombination substrate were used for monitoring somatic homologous recombination frequency (HRF). The recombination substrate consists of a non-functional chimaeric *uidA* (*GUS*) gene containing partially overlapping homologous sequence (Fig. 1a). Lines IC1 and IC9 carry an intermolecular recombination substrate; lines 1406, 1445, 11 and 651 carry an intramolecular recombination substrate. Wild-type *A. thaliana* plants (ecotype Columbia) were also used for crosses. S₁ plants are defined as the progeny of either untreated or treated S₀ plants. S₂, S₃ and S₄ plants correspond to the subsequent generations.

Growth conditions. For soil-cultured plants, seeds were sown (2 per pot) and put at 4 °C in the dark for 3 days. The pots were transferred to a phytotron (70% humidity) and kept under a cycle of 16-h-light (20 °C) and 8-h-dark (16 °C). For *in vitro* culture, plants were germinated on GM medium (MS salts (Duchefa), 1% sucrose, 0.8% agar, pH 5.8). Plants were grown in a culture chamber under a 16-h-light (20 °C) and 8-h-dark (16 °C) photoperiod. Three-week-old plants were transferred to soil and grown for seed production (self-pollination) or crosses.

Treatment with UV-C and flagellin. For all experiments, plants were grown *in vitro* on solid GM medium (MS salts (Duchefa), 1% sucrose, 0.8% Agar-agar ultrapure (Merck), pH 5.8) in a culture chamber under a 16-h-light (20 °C) and 8-h-dark (16 °C) photoperiod for 13 days before being subjected to the specific preculture condition established for each treatment. For UV-C treatment, plants were transferred to large Petri dishes (160-mm diameter) containing solid GM medium, to a density of 1 plant per cm². For the UV-C irradiation (254 nm, 6 kerg cm⁻²), a Mineral light-lamp (UV-Products) was used. For the flagellin treatment with either the active peptide (flg22) or the inactive peptide (flg22 *A. tum.*), plants were subcultured for 24 h in 300 µl of liquid GM medium in 24-cell plates. Flg22 or flg22 *A. tum.* (1 µM) diluted in liquid GM medium was applied. Four days after each treatment, plants were transferred to soil and grown for seed production (self-pollination) or crosses.

Monitoring of somatic homologous recombination events. For monitoring the somatic homologous recombination events, the histochemical GUS assay²⁶ was performed on 3-week-old *in vitro* germinated plants. The HRF corresponds to the average number of blue spots (equivalent to the number of homologous recombination events) per plant $\pm$ s.e.m., in a population. For each recombination line, experiments were at least duplicated. The *t*-test was used for statistical analyses with *P* = 0.05. Note that somatic HRF can only be compared within experimental series due to the specific culture conditions used for different treatments: solid culture medium and liquid culture medium for UV-C and flg22, respectively.

Received 29 March; accepted 29 June 2006.

Published online 6 August 2006.

1. McClintock, B. The significance of responses of the genome to challenge. *Science* **226**, 792–801 (1984).

2. Ries, G. *et al.* Elevated UV-B radiation reduces genome stability in plants. *Nature* **406**, 98–101 (2000).
3. Lucht, J. M. *et al.* Pathogen stress increases somatic recombination frequency in *Arabidopsis*. *Nature Genet.* **30**, 311–314 (2002).
4. Kovalchuk, I. *et al.* Pathogen-induced systemic plant signal triggers DNA rearrangements. *Nature* **423**, 760–762 (2003).
5. Cullis, C. A. Mechanisms and control of rapid genomic changes in flax. *Ann. Bot. (Lond.)* **95**, 201–206 (2005).
6. Zipfel, C. *et al.* Bacterial disease resistance in *Arabidopsis* through flagellin perception. *Nature* **428**, 764–767 (2004).
7. Mottinger, J. P., Johns, M. A. & Freeling, M. Mutations of the *Adh1* gene in maize following infection with barley stripe mosaic virus. *Mol. Gen. Genet.* **195**, 367–369 (1984).
8. Walbot, V. Reactivation of Mutator transposable elements of maize by ultraviolet light. *Mol. Gen. Genet.* **234**, 353–360 (1992).
9. Grandbastien, M. A. *et al.* Stress activation and genomic impact of Tnt1 retrotransposons in Solanaceae. *Cytogenet. Genome Res.* **110**, 229–241 (2005).
10. Kovalchuk, I., Kovalchuk, O. & Hohn, B. Genome-wide variation of the somatic mutation frequency in transgenic plants. *EMBO J.* **19**, 4431–4438 (2000).
11. Lebel, E. G., Masson, J., Bogucki, A. & Paszkowski, J. Stress-induced intrachromosomal recombination in plant somatic cells. *Proc. Natl Acad. Sci. USA* **90**, 422–426 (1993).
12. Walbot, V. Sources and consequences of phenotypic plasticity in flowering plants. *Trends Plant Sci.* **1**, 27–32 (1996).
13. Swoboda, P., Gal, S., Hohn, B. & Puchta, H. Intrachromosomal homologous recombination in whole plants. *EMBO J.* **13**, 484–489 (1994).
14. Molinier, J., Ries, G., Bonhoeffer, S. & Hohn, B. Interchromatid and interhomolog recombination in *Arabidopsis thaliana*. *Plant Cell* **16**, 342–352 (2004).
15. Schuermann, D., Molinier, J., Fritsch, O. & Hohn, B. The dual nature of homologous recombination in plants. *Trends Genet.* **21**, 172–181 (2005).
16. Felix, G., Duran, J. D., Volko, S. & Boller, T. Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. *Plant J.* **18**, 265–276 (1999).
17. Klumpp, A., Ansel, W., Fomin, A., Schnirring, S. & Pickl, C. Influence of climatic conditions on the mutations in pollen mother cells of *Tradescantia* clone 4430 and implications for the Trad-MCN bioassay protocol. *Hereditas* **141**, 142–148 (2004).
18. Puchta, H., Swoboda, P. & Hohn, B. Induction of intrachromosomal homologous recombination in whole plants. *Plant J.* **7**, 203–210 (1995).
19. Jorgensen, R. A. Restructuring the genome in response to adaptive challenge: McClintock's bold conjecture revisited. *Cold Spring Harb. Symp. Quant. Biol.* **69**, 349–354 (2004).
20. Madlung, A. & Comai, L. The effect of stress on genome regulation and structure. *Ann. Bot. (Lond.)* **94**, 481–495 (2004).
21. Cubas, P., Vincent, C. & Coen, E. An epigenetic mutation responsible for natural variation in floral symmetry. *Nature* **401**, 157–161 (1999).
22. Jacobsen, S. & Meyerowitz, E. M. Hypermethylated *SUPERMAN* epigenetic alleles in *Arabidopsis*. *Science* **277**, 1100–1103 (1997).
23. Jablonka, E. & Lamb, M. J. *Epigenetic Inheritance and Evolution: The Lamarckian Dimension* (Oxford Univ. Press, Oxford, 1995).
24. Gherbi, H. *et al.* Homologous recombination in *planta* is stimulated in the absence of Rad50. *EMBO Rep.* **2**, 287–291 (2001).
25. Puchta, H., Swoboda, P., Gal, S., Blot, M. & Hohn, B. Somatic intrachromosomal recombination events in populations of plant siblings. *Plant Mol. Biol.* **28**, 281–292 (1995).
26. Jefferson, R. A. Assaying chimeric genes in plants: the GUS gene fusion system. *Plant Mol. Biol. Rep.* **5**, 387–405 (1987).

Acknowledgements We acknowledge the critical analysis of the manuscript by O. M. Scheid, D. Schuermann, R. Jorgensen, U. Grossniklaus, D. Schuebeler, L. Comai and T. Boller. We are grateful to the Novartis Research Foundation and the European Union project PLANTREC for financial support.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.H. (barbara.hohn@fmi.ch).

LETTERS

Dynamics of heat shock factor association with native gene loci in living cells

Jie Yao^{1,2}, Katherine M. Munson³, Watt W. Webb^{1,2} & John T. Lis^{1,3}

Direct observation of transcription factor action in the living cell nucleus can provide important insights into gene regulatory mechanisms^{1,2}. Live-cell imaging techniques have enabled the visualization of a variety of intranuclear activities, from chromosome dynamics³ to gene expression⁴. However, progress in studying transcription regulation of specific native genes has been limited, primarily as a result of difficulties in resolving individual gene loci and in detecting the small number of protein molecules functioning within active transcription units. Here we report that multiphoton microscopy imaging⁵ of polytene nuclei in living *Drosophila* salivary glands allows real-time analysis of transcription factor recruitment and exchange on specific native genes. After heat shock, we have visualized the recruitment of RNA polymerase II (Pol II) to native *hsp70* gene loci 87A and 87C in real time. We show that heat shock factor (HSF), the transcription activator of *hsp70*, is localized to the nucleus before heat shock and translocates from nucleoplasm to chromosomal loci after heat shock. Assays based on fluorescence recovery after photobleaching⁶ show a rapid exchange of HSF at chromosomal loci under non-heat-shock conditions but a very slow exchange after heat shock. However, this is not a consequence of a change of HSF diffusibility, as shown here directly by fluorescence correlation spectroscopy⁷. Our results provide strong evidence that activated HSF is stably bound to DNA *in vivo* and that turnover or disassembly of transcription activator is not required for rounds of *hsp70* transcription. This and previous studies^{8,9} indicate that transcription activators display diverse dynamic behaviours in their associations with targeted loci in living cells. Our method can be applied to study the dynamics of many factors involved in transcription and RNA processing, and in their regulation at native heat shock genes *in vivo*.

Drosophila polytene nuclei contain extended, interphase-like giant chromosomes produced by endoreduplication processes. Polytene chromosomes have been invaluable in providing the first links between genetic and physical maps and in allowing the mapping of cloned sequences¹⁰. Additionally, they have provided a method of examining the distribution of specific proteins on chromosomal sites¹¹ and have served as a structural model of chromosome organization within the interphase nucleus¹². Here we describe and apply a method for the imaging of *Drosophila* polytene nuclei in living salivary gland tissue with the use of multiphoton microscopy (MPM)⁵, which has exceptional optical-sectioning capability in thick specimens¹³. The MPM optical sections of DNA-stained, cultured salivary gland show a clear chromosome banding pattern in which chromosomal loci can easily be recognized (Fig. 1a, b, and Supplementary Fig. S1, the complete section series). As noted previously¹², the centromeric region and the telomeres are found at opposite ends of the nucleus, indicating the Rabl orientation of

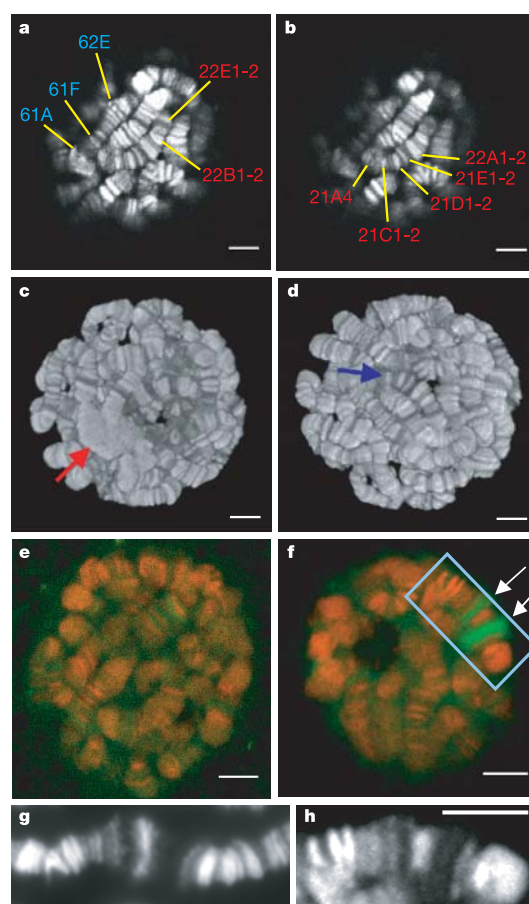


Figure 1 | Multiphoton imaging of *Drosophila* polytene nuclei in living cells. **a, b**, Two consecutive optical sections of a polytene nucleus stained with Hoechst33342. The *z*-distance is 0.5 μ m. Recognized bands on chromosome arm 2L (red) and 3L (blue) are noted. **c, d**, The Arctic (**c**) and Antarctic (**d**) views of a three-dimensional reconstructed polytene nucleus. Red and blue arrows indicate the centromeric region (chromocentre) and telomere, respectively. **e, f**, Optical sections of polytene nuclei containing EGFP-Rpb3 (green) co-stained with Hoechst (red) (**e**) at room temperature and (**f**) after 20 min of HS at 36.5 °C. Major HS puffs 87A and 87C are indicated by the white arrows, with the neighbouring region in a cyan box shown in **h, g**. A spread Hoechst-stained chromosome showing the 87A and 87C heat shock puffs. **h**, Enlarged image of major HS puff 87A and 87C (Hoechst-stained) region from **f**. Scale bars, 5 μ m.

¹Field of Biochemistry, Molecular and Cell Biology, ²School of Applied and Engineering Physics and ³Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853, USA.

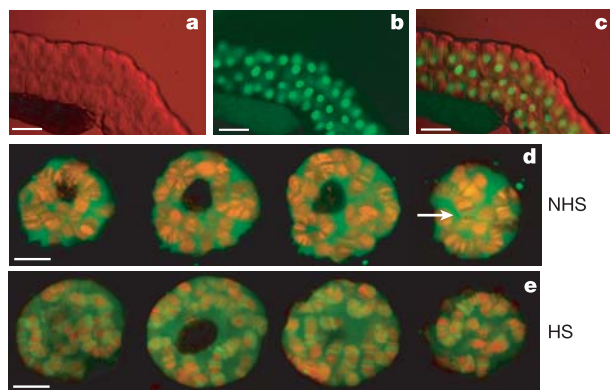


Figure 2 | Expression and localization of HSF-EGFP, and translocation of HSF-EGFP from nucleoplasm to chromosomal loci upon HS. **a–c**, Nuclear localization of HSF-EGFP by imaging live salivary gland tissue in culture. **a**, Differential interference contrast image. **b**, Wide-field epifluorescence image of HSF-EGFP. **c**, Merge of **a** and **b**. Scale bars, 100 μ m. **d**, **e**, Two-photon optical sections of polytene nuclei expressing HSF-EGFP (green) and stained with Hoechst33342 (red), when the salivary gland was cultured at room temperature (**d**) or at 36.5 $^{\circ}$ C (**e**) for 30 min. Scale bars, 10 μ m.

chromosomes (Fig. 1c, d). The image quality is significantly improved over the previous results with wide-field microscopy¹². MPM imaging of the *Drosophila* polytene nucleus therefore provides unprecedented effectiveness in resolving individual native gene loci.

Green fluorescent proteins (GFPs) have been widely used to study protein localization and dynamics in living cells¹⁴. To follow the transcriptional activation of genes, we generated a fly line expressing enhanced green fluorescent protein (EGFP)-tagged Pol II in larval salivary glands. The tagged subunit Rpb3 is an essential and highly conserved subunit of Pol II and is homologous to the α -subunit of *Escherichia coli* RNA polymerase¹⁵. Examining the spread polytene chromosomes stained with anti-GFP and with either anti-Rpb3 or the antibody against the largest subunit of Pol II (Rpb1) shows that the resulting patterns overlap at more than 150 sites (Supplementary Fig. S2; see also ref. 16). Under non-heat-shock (NHS) conditions, EGFP-Rpb3 is broadly distributed at interbands and puffs (Fig. 1e); after heat shock (HS), EGFP-Rpb3 redistributes and becomes highly enriched at the heat shock puffs, including the prominent puff doublet at the 87A and 87C loci, which contains the *hsp70* genes

(Fig. 1f–h). These results are perfectly congruent with previous antibody staining of fixed polytene chromosomes¹⁶ and with higher-resolution biochemical studies of heat shock gene expression after HS¹⁷. Our results indicate that EGFP-Rpb3 has been incorporated into the Pol II transcription machinery and thus that it is a valid indicator of functional Pol II. Furthermore, we have recorded the time series of EGFP-Rpb3 recruitment to these heat shock loci after HS (Supplementary Fig. S3a) and demonstrated that the observed doublet structure is indeed heat-induced and can be unambiguously assigned to 87A and 87C loci. We found that Pol II recruitment accompanies the puff growth and reaches a plateau in about 10 min after full HS (Supplementary Fig. S3b). MPM imaging of the *Drosophila* polytene nucleus can therefore be used to study the recruitment and dynamics of particular transcription factors at specific native genes.

A particularly important set of transcription factors are those that drive the activation of specific sets of genes by binding to the regulatory elements associated with these targeted genes. Once bound, these activators interact with other coactivators to stimulate gene expression. Many of the DNA and protein targets of these activators have been identified^{18,19}; however, the molecular dynamics of these factors during the activation and maintenance of transcription *in vivo* remains poorly understood. HSF is a strong transcription activator with an acidic activation domain²⁰. It is generally agreed that heat shock triggers the trimerization of HSF, binding to specific DNA sequences, the heat shock elements (HSEs), and the activation of heat shock gene transcription²⁰; however, controversy remains over whether heat shock triggers nuclear import of HSF²¹, or whether HSF normally resides in the nucleus²².

To reveal HSF dynamics in living cells, we have generated a fly line expressing EGFP-tagged HSF at the carboxy terminus and have shown that this HSF-EGFP has a localization pattern on polytene chromosomes indistinguishable from that of native HSF (Supplementary Fig. S4). A western blot with the HSF antibody shows that the HSF-EGFP is expressed at a level approximately equal to the endogenous HSF (Supplementary Fig. S5) and should not significantly perturb the normal HSF equilibrium. Before HS, HSF is localized in the nuclei of polytene cells (Fig. 2a–c). Within the nucleus, HSF is mostly located in the nucleoplasm but is also detectable at several loci, the most prominent being at the 63B locus²² (Fig. 2d; see Supplementary Fig. S6 for chromosome mapping). After HS, HSF is rapidly recruited from the nucleoplasm to many native loci, including the heat shock genes (Fig. 2e, and

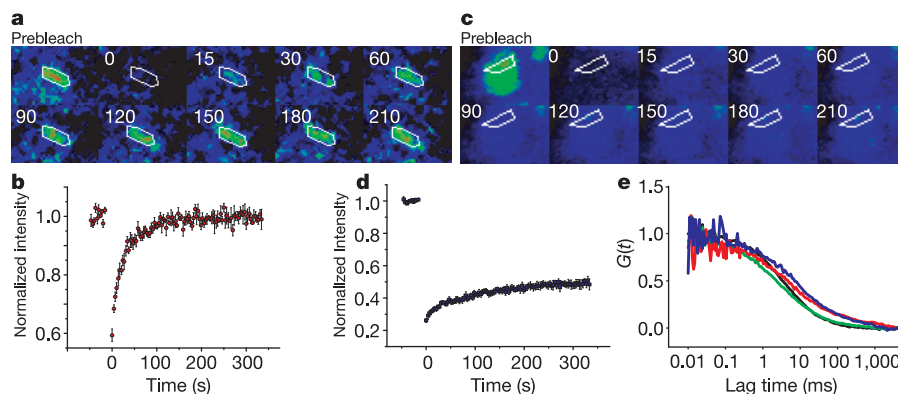


Figure 3 | Distinct exchange kinetics of HSF at specific chromosomal loci during NHS and HS conditions. **a**, **b**, During NHS. **a**, Images of the bleached region before bleaching and the time course afterwards. Time after bleaching in seconds is shown on each image. The strongest HSF binding site is bordered by a white line. **b**, The normalized fluorescence intensities of the HSF-binding site (within the white line) before and during FRAP recovery; $n = 4$. **c**, **d**, During HS. **c**, The bleached region during FRAP (87A site; see Supplementary Fig. S7). Time after bleaching in seconds is shown on each

image. **d**, FRAP recovery intensity plots. The normalized intensities are the average of numerous independent experiments (shown in Supplementary Fig. S8); $n = 5$. **e**, Diffusion of the nucleoplasmic HSF-EGFP measured by FCS at NHS (red) and HS (blue). FCS curves of free GFP in the polytene nucleus are also shown (NHS, black; HS, green). $G(t)$ values are normalized to $G(0) = 1$ for these comparative data. Where shown, error bars indicate s.e.m.

Supplementary Movie S1). Therefore, in agreement with previous studies^{20,22}, the heat-induced binding of HSF to specific chromosomal sites in *Drosophila* is not a consequence of nuclear import but can instead be explained simply by the heat-induced association of nucleoplasmic monomeric HSF to form trimers, which have a higher affinity than monomers for HSEs.

Once bound to their DNA target, some transcription activators, such as the glucocorticoid receptor (GR), have been shown to exchange rapidly between the nucleoplasm and a synthetic tandem array of its DNA response elements in living cells⁸. High mobility and transient binding have been proposed to be general features of DNA regulatory proteins⁹. Active turnover mechanisms have also been suggested to ensure the binding of fresh transcription activators to start new rounds of transcription^{23,24}. However, the exchange dynamics needs to be examined on native genes and with other strong and regulated activators such as HSF. In polytene nuclei, the individual native chromosomal loci to which HSF binds can be examined directly by MPM imaging.

We used fluorescence recovery after photobleaching (FRAP) to study the HSF exchange at native gene loci on a timescale from seconds to minutes. Before HS, HSF is localized at a few chromosomal loci; the major association site is the 63B locus²² (Fig. 2d, arrow). Photobleaching of HSF-EGFP at the 63B locus is followed by a rapid recovery with a half-recovery time of about 15 s, which indicates a rapid exchange as also seen for GR^{8,25} (Fig. 3a, b). At 10 min after HS, the major HS puff can be found by locating the adjacent pair of puff sites containing the most prominent levels of HSF (Fig. 3d; see Supplementary Fig. S7 for the complete z-series of the same nucleus, which identifies this unique pair of HS puffs). In sharp contrast to the rapid recovery seen in NHS, very slow, if any, recovery is found ($t_{1/2} > 6$ min) at these chromosome loci (Fig. 3c, d). The slow recovery was found for various bleached fractions (that is, changing bleaching energy; see Supplementary Fig. S8 and Supplementary Discussion) and is also found at other unpuffed HSF-binding sites, which are much less transcriptionally active or inactive during heat shock (data not shown). Moreover, fluorescence correlation spectroscopy (FCS)⁷ on the nucleoplasmic HSF-EGFP during NHS and HS shows very similar curves (Fig. 3e), indicating that the diffusion mobility of HSF is only slightly altered after HS, as can be explained by the trimerization of HSF. The slow recovery shown in FRAP therefore must indicate stable binding of HSF to DNA during HS.

The distinctively different recovery patterns of HSF under NHS and under HS corresponds *in vivo* to the marked difference in the DNA-binding affinity of HSF monomers (NHS) and trimers (HS)²⁰. We therefore propose that a transcription activator's exchange dynamics on its targets may simply reflect the dissociation rate constant of the protein-promoter complex. The low affinity of some activators leads to their transient binding and has been suggested to cause the probabilistic assembly of transcriptional machinery⁹. The high affinity of other activators leads to their stable binding as we have seen, and this in turn is conducive to the formation of stable coactivator assemblies and the efficient recruitment of Pol II for repeated cycles of transcription. The exchange dynamics of some activators may involve other mechanisms; for instance, NF- κ B, which has high affinity for DNA, was found to exchange rapidly at the tandem-repeat target gene loci²⁶. In addition, chromatin remodelling might have a function in these processes (Supplementary Discussion).

The slow exchange of activated HSF at the *hsp70* promoter presents a sharp contrast with the rapid recruitment and elongation of RNA polymerase II at *hsp70* genes during HS. During a 2-min transcription cycle (that is, the time it takes a Pol II molecule to transcribe the *hsp70* gene^{17,27}), more than 20 Pol II molecules have begun the transcription of each *hsp70* gene¹⁷; however, very little new HSF has bound to the gene as shown by FRAP (Fig. 3c, d). Therefore, our data do not support the 'activation by destruction' hypothesis

that the recruitment of new polymerase requires the ubiquitin-proteasome system (UPS) to turn over the 'spent' activator on the promoter²⁴. Moreover, more than the total amount of intracellular HSF would be degraded during a short period of heat shock if 'activation by destruction' were true for every round of heat shock gene transcription (Supplementary Discussion). HSF is an acidic, strong activator, like many positive regulatory factors, and *hsp70* transcription resembles that of many other genes¹⁷. Recent results²⁸ on the yeast Gal4 activator have shown that it, too, is stably bound to its regulatory sites during gene activation. Therefore two independent and complementary approaches on the two widely studied acidic activators have revealed their stable binding to DNA during gene activation. Alternative models for activator function that propose activator recycling as a key component, such as hit and run⁸, chaperone-assisted disassembly²⁹ or UPS-mediated turnover²⁴, can apply to some but clearly not all transcriptional activators.

The stable binding of HS-activated HSF and the transient binding of ligand-activated GR collectively show the diverse 'action modes' of transcription activators: both stably bound and transiently bound activators can support gene transcription. How individual activators function in these two modes on their respective gene targets remains to be seen, with the underlying mechanisms yet to be determined. Importantly, the dynamic behaviour of coactivators, Pol II transcription and RNA-processing machinery at native mRNA genes is largely unknown in living cells, and the described experimental approach will be applicable to further investigations.

METHODS

Fly strains and crosses. The fly lines expressing GFP-tagged Rpb3 and HSF were created by transforming *Drosophila* germline cells (Genetic Services) using the Gateway P-element insertion vectors (<http://www.ciwemb.edu/labs/murphy/Gateway%20vectors.html>) and crossed with Gal4 driver lines expressing the GFP-tagged proteins in salivary glands. The fly lines used were the following: line 1, EGFP-Rpb3: P[w⁺, UAS-EGFP-Rpb3], P[w⁺, C147-Gal4]; line 2, HSF-EGFP: P[w⁺, UAS-HSF-EGFP], P[w⁺, Sgs3-Gal4], and line 3, nuclear-localized GFP: P[w⁺, C147-Gal4]/+; P[w⁺, UAS-NLS-NES^{P12}-GFP]/+. Lines 1 and 2 are homozygous and have the two P-element insertions recombined on the same chromosome (line 1, chromosome 2; line 2, chromosome 3). Line 3 was made by crossing the P-element strains containing the corresponding UAS-GFP and Gal4 alleles. All Gal4 driver lines and the UAS-GFP line were obtained from the Bloomington *Drosophila* Stock Center.

MPM imaging, FRAP and FCS of polytene nuclei. Multiphoton microscopy of polytene nuclei was performed with a laboratory-built Bio-Rad laser scanning unit (MRC 1024) coupled to an Olympus IX-70 inverted microscope with a Carl Zeiss 40X F-FLUAR objective (numerical aperture 1.3, oil immersion). The excitation was from a Ti:sapphire mode-locked laser (Tsunami; Spectra Physics) providing approximately 100-fs pulses at an 80-MHz repetition rate. Intact, healthy salivary glands were chosen and imaged within an FCS2 closed chamber system (Bioptechs) for the inverted microscope. Chamber temperature was controlled at room temperature (about 22 °C for NHS) or shifted to 36.5 °C for inducing HS. Polytene nuclei were usually imaged at a depth of 30–100 μ m into the salivary gland tissue. Three-dimensional reconstruction of the Hoechst-stained optical sections of a polytene nucleus was performed by Vox software (<http://www.nephrology.iupui.edu/imaging/voxx/>). FRAP and FCS studies in polytene nuclei were performed in a custom-built MPM imaging system described in ref. 30. EGFP fluorescence was directed through an HQ575/150m emission filter (Chroma) and was detected by a GaAsP photomultiplier tube (Hamamatsu Photonics). FRAP of HSF-EGFP at chromosomal sites was performed by repetitive raster-scanning these sites at an excitation power of about 32 mW followed by time-lapse imaging of the same section every 3 s for up to 10 min. During FCS recording, the laser beam was stationed within the nucleoplasmic region of the polytene nucleus. The excitation power was less than 3 mW. Each FCS curve was the average of ten 30-s runs. Detailed experimental procedures are available in Supplementary Information.

Received 19 March; accepted 29 June 2006.

Published online 23 August 2006.

1. Spector, D. L. The dynamics of chromosome organization and gene regulation. *Annu. Rev. Biochem.* **72**, 573–608 (2003).
2. Shav-Tal, Y., Singer, R. H. & Darzacq, X. Imaging gene expression in single living cells. *Nature Rev. Mol. Cell Biol.* **5**, 855–861 (2004).

3. Belmont, A. S. Visualizing chromosome dynamics with GFP. *Trends Cell Biol.* **11**, 250–257 (2001).
4. Janicki, S. M. *et al.* From silencing to gene expression: real-time analysis in single cells. *Cell* **116**, 683–698 (2004).
5. Denk, W., Strickler, J. H. & Webb, W. W. Two-photon laser scanning fluorescence microscopy. *Science* **248**, 73–76 (1990).
6. Axelrod, D., Koppel, D. E., Schlessinger, J., Elson, E. & Webb, W. W. Mobility measurement by analysis of fluorescence photobleaching recovery kinetics. *Biophys. J.* **16**, 1055–1069 (1976).
7. Magde, D., Elson, E. & Webb, W. W. Thermodynamic fluctuations in a reacting system—measurement by fluorescence correlation spectroscopy. *Phys. Rev. Lett.* **29**, 705–708 (1972).
8. McNally, J. G., Muller, W. G., Walker, D., Wolford, R. & Hager, G. L. The glucocorticoid receptor: rapid exchange with regulatory sites in living cells. *Science* **287**, 1262–1265 (2000).
9. Misteli, T. Protein dynamics: implications for nuclear architecture and gene expression. *Science* **291**, 843–847 (2001).
10. Bridges, C. B. Salivary chromosome maps. *J. Hered.* **26**, 60–64 (1935).
11. Schwartz, B. E., Werner, J. K. & Lis, J. T. Indirect immunofluorescent labeling of *Drosophila* polytene chromosomes: visualizing protein interactions with chromatin *in vivo*. *Methods Enzymol.* **376**, 393–404 (2004).
12. Agard, D. A. & Sedat, J. W. Three-dimensional architecture of a polytene nucleus. *Nature* **302**, 676–681 (1983).
13. Zipfel, W. R., Williams, R. M. & Webb, W. W. Nonlinear magic: multiphoton microscopy in the biosciences. *Nature Biotechnol.* **21**, 1369–1377 (2003).
14. Zhang, J., Campbell, R. E., Ting, A. Y. & Tsien, R. Y. Creating new fluorescent probes for cell biology. *Nature Rev. Mol. Cell Biol.* **3**, 906–918 (2002).
15. Cramer, P. Multisubunit RNA polymerases. *Curr. Opin. Struct. Biol.* **12**, 89–97 (2002).
16. Weeks, J. R., Hardin, S. E., Shen, J., Lee, J. M. & Greenleaf, A. L. Locus-specific variation in phosphorylation state of RNA polymerase II *in vivo*: correlations with gene activity and transcript processing. *Genes Dev.* **7**, 2329–2344 (1993).
17. Lis, J. Promoter-associated pausing in promoter architecture and postinitiation transcriptional regulation. *Cold Spring Harb. Symp. Quant. Biol.* **63**, 347–356 (1998).
18. Green, M. R. Eukaryotic transcription activation: right on target. *Mol. Cell* **18**, 399–402 (2005).
19. Reeves, W. M. & Hahn, S. Targets of the Gal4 transcription activator in functional transcription complexes. *Mol. Cell Biol.* **25**, 9092–9102 (2005).
20. Wu, C. Heat shock transcription factors: structure and regulation. *Annu. Rev. Cell Dev. Biol.* **11**, 441–469 (1995).
21. Zandi, E., Tran, T. N., Chamberlain, W. & Parker, C. S. Nuclear entry, oligomerization, and DNA binding of the *Drosophila* heat shock transcription factor are regulated by a unique nuclear localization sequence. *Genes Dev.* **11**, 1299–1314 (1997).
22. Westwood, J. T., Clos, J. & Wu, C. Stress-induced oligomerization and chromosomal relocalization of heat-shock factor. *Nature* **353**, 822–827 (1991).
23. Muratani, M., Kung, C., Shokat, K. M. & Tansey, W. P. The F box protein Dsg1/Mdm30 is a transcriptional coactivator that stimulates Gal4 turnover and cotranscriptional mRNA processing. *Cell* **120**, 887–899 (2005).
24. Lipford, J. R., Smith, G. T., Chi, Y. & Deshaies, R. J. A putative stimulatory role for activator turnover in gene expression. *Nature* **438**, 113–116 (2005).
25. Stavreva, D. A., Muller, W. G., Hager, G. L., Smith, C. L. & McNally, J. G. Rapid glucocorticoid receptor exchange at a promoter is coupled to transcription and regulated by chaperones and proteasomes. *Mol. Cell Biol.* **24**, 2682–2697 (2004).
26. Bosisio, D. *et al.* A hyper-dynamic equilibrium between promoter-bound and nucleoplasmic dimers controls NF- κ B-dependent gene activity. *EMBO J.* **25**, 798–810 (2006).
27. O'Brien, T. & Lis, J. T. Rapid changes in *Drosophila* transcription after an instantaneous heat shock. *Mol. Cell Biol.* **13**, 3456–3463 (1993).
28. Nalley, K., Johnston, S. A. & Kodadek, T. Proteolytic turnover of the Gal4 transcription factor is not required for function *in vivo*. *Nature* doi:10.1038/nature05067 (this issue).
29. Freeman, B. C. & Yamamoto, K. R. Disassembly of transcriptional regulatory complexes by molecular chaperones. *Science* **296**, 2232–2235 (2002).
30. Larson, D. R., Ma, Y. M., Vogt, V. M. & Webb, W. W. Direct measurement of Gag–Gag interaction during retrovirus assembly with FRET and fluorescence correlation spectroscopy. *J. Cell Biol.* **162**, 1233–1244 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature

Acknowledgements We thank J. K. Werner for immunofluorescence labeling of polytene chromosomes; K. L. Zobeck for initial western blot experiments; R. M. Williams and W. R. Zipfel for technical assistance with MPM imaging; T. Kodadek for exchanging manuscripts before submission; and W. L. Kraus for critical reading of the manuscript. This work was performed partly in the Developmental Resources for Biophysical Imaging Opto-Electronics and was supported by an NSF grant to W.W.W. and J.T.L., an NIH grant to J.T.L., and an NIH-NIBIB grant to W.W.W.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.T.L. (jtl10@cornell.edu) or W.W.W. (www2@cornell.edu).

LETTERS

Proteolytic turnover of the Gal4 transcription factor is not required for function *in vivo*

Kip Nalley¹, Stephen Albert Johnston^{1†} & Thomas Kodadek¹

Transactivator–promoter complexes are essential intermediates in the activation of eukaryotic gene expression. Recent studies of these complexes have shown that some are quite dynamic in living cells¹ owing to rapid and reversible disruption of activator–promoter complexes by molecular chaperones^{2–6}, or a slower, ubiquitin–proteasome–pathway-mediated turnover of DNA-bound activator^{7–9}. These mechanisms may act to ensure continued responsiveness of activators to signalling cascades by limiting the lifetime of the active protein–DNA complex. Furthermore, the potency of some activators is compromised by proteasome inhibition, leading to the suggestion that periodic clearance of activators from a promoter is essential for high-level expression^{8,10–12}. Here we describe a variant of the chromatin immunoprecipitation assay that has allowed direct observation of the kinetic stability of native Gal4–promoter complexes in yeast. Under non-inducing conditions, the complex is dynamic, but on induction the Gal4–promoter complexes ‘lock in’ and exhibit long half-lives. Inhibition of proteasome-mediated proteolysis had little or no effect on Gal4-mediated gene expression. These studies, combined with earlier data, show that the lifetimes of different transactivator–promoter complexes *in vivo* can vary widely and that proteasome-mediated turnover is not a general requirement for transactivator function.

To study the dynamics of Gal4–promoter complexes in living yeast cells we developed the competition assay shown schematically in Fig. 1a. The method relies on the fact that fusion proteins containing the ligand binding domain (LBD) of nuclear hormone receptors such as oestrogen receptor- α (ER- α) are inactive due to high-affinity interactions with Hsp90 (refs 13, 14). We expressed in yeast an Myc–Gal4(DBD)–ER(LBD)–VP16–Flag fusion protein (Myc–G4–ER–VP16) from the powerful *ADH1* promoter¹⁵, providing a high concentration of the Hsp90-masked protein. Addition of β -oestradiol, which is known to disrupt the ER(LBD)–Hsp90 interaction¹⁶, should result in the sudden release of a large excess of DNA-binding-competent Myc–G4–ER–VP16 (Supplementary Fig. 1). Therefore, dissociation of native Gal4 from a promoter would result in irreversible loss of the complex due to competition with the large excess of the LBD-containing fusion protein. Using antibodies that differentiate the proteins, one could then use time-resolved chromatin immunoprecipitation (ChIP) assays to monitor the net rate of transcription factor exchange after steroid addition.

To validate this idea, a $\Delta gal4$ strain was transformed with a plasmid expressing Myc–G4–ER–VP16 under the control of the *ADH1* promoter. ChIP analysis using an anti-Myc polyclonal antibody failed to reveal the presence of the engineered protein on the *GAL1/10* promoter before addition of steroid (Fig. 1b), validating that the apo-protein is unable to access the promoter. On treatment with β -oestradiol, the Myc–G4–ER–VP16 protein could be detected

on the promoter, by ChIP using anti-Myc antibody, within 1–5 min. The signal reaches a maximum by 15 min after steroid addition. Myc–G4–ER–VP16 did not associate significantly with the closely related *PUT2* promoter¹⁷ (Supplementary Fig. 2), confirming that the incoming competitor protein binds specifically to *GAL* promoters. Finally, a robust increase in *GAL1*, but not *PUT2*, transcription was seen only after steroid addition (Fig. 1c and Supplementary Fig. 2).

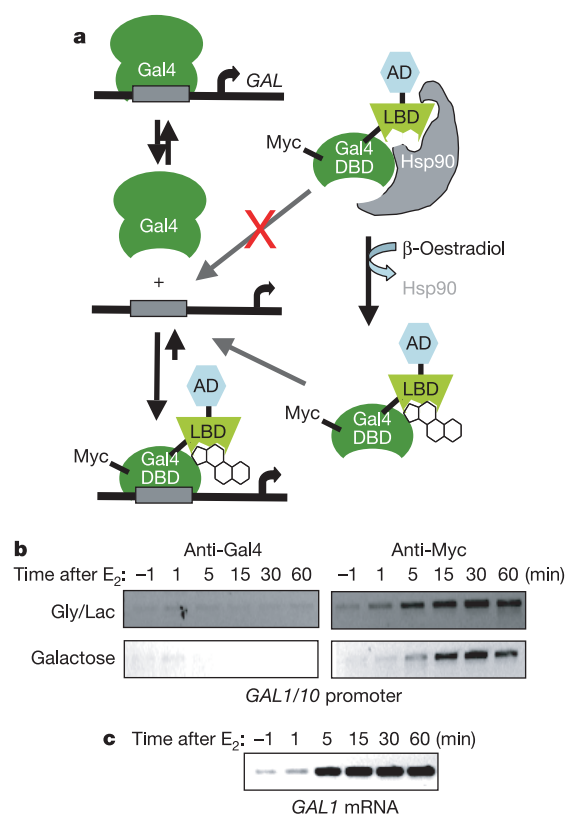


Figure 1 | The competition ChIP assay as applied to the yeast transcription factor Gal4. **a**, Schematic representation of the assay. See text for details. **b**, ChIP assays performed in $\Delta gal4$ cells expressing the Myc–Gal4(DBD)–ER(LBD)–VP16–Flag (Myc–G4–ER–VP16) fusion protein. Cells were grown in minimal medium with glycerol/lactic acid or galactose as the carbon source. β -Oestradiol (E_2) was added at $t = 0$. The -1 time point was taken 1 min before addition of β -oestradiol. Myc–G4–ER–VP16 binding to chromatin was probed by chromatin immunoprecipitation using anti-Myc agarose and amplifying the region of interest by PCR. **c**, Transcript levels for the *GAL1* gene measured by RT–PCR before or after addition of β -oestradiol (times indicated) in $\Delta gal4$ cells expressing Myc–G4–ER–VP16.

¹Departments of Internal Medicine and Molecular Biology, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, Texas 75390-9185, USA.

[†]Present address: Center for Innovations in Medicine, Biodesign Institute, Arizona State University, 1001 S. McAllister Ave, Tempe, Arizona 85287-5001, USA.

Taken together, these data show that Myc-G4-ER-VP16 is inactive until addition of steroid, at which point it associates rapidly with GAL promoters.

Native Gal4 binds to its cognate promoters but is prevented from driving transcription in non-inducing media (glycerol plus lactic acid) due to masking of the activation domain by the specific repressor Gal80 (ref. 18). On addition of galactose (inducing conditions), Gal80 dissociates rapidly and Gal4 potently drives expression of its target genes. Yeast expressing Myc-G4-ER-VP16 were grown in non-inducing media then induced with galactose. As expected, Gal4-GAL1/10 promoter complexes can be precipitated with an anti-Gal4 polyclonal antibody raised against the carboxy terminus of the Gal4 protein before the addition of β -oestradiol (Fig. 2a). Notably, this signal decreased only slightly over the course of 1 h after the addition of steroid (Fig. 2a and Supplementary Fig. 3a), despite the presence of a large excess of free Myc-G4-ER-VP16 competitor protein. Correspondingly, there was only a small amount of the competitor fusion protein bound to the GAL1/10 promoter, as demonstrated by the weak signal obtained using the anti-Myc antibody (Fig. 2a). This suggests that the native Gal4-promoter complex is quite stable on the GAL1/10 promoter, with a half-life of approximately 1 h. This experiment was repeated with different yeast strains and different antibodies in the ChIP assay, and in all cases, the Gal4-UAS complexes were found to be similarly stable. Indeed, the apparent half-life of the active Gal4-GAL1/10 promoter complex is similar to the doubling time of the strain used (Supplementary Fig. 4), suggesting that replication may be required to disrupt this complex, although this hypothesis has not been tested directly. Reporter gene experiments confirmed that over this period massive induction of GAL gene expression occurred (data not shown).

The GAL1/10 promoter contains three strong and one weak Gal4-binding sites. Therefore, we also examined the stability of the Gal4-GAL3 promoter complex, because the GAL3 promoter contains only a single activator-binding site. This complex was also quite stable, although some dissociation occurred over the course of an hour and, correspondingly, some binding of the LBD-containing competitor

was detected (Fig. 2a). These data show that even in the absence of cooperative binding, Gal4 dissociates slowly under inducing conditions. The even slower dissociation of the multiple Gal4 dimers from GAL1/10 suggests further stabilization due to cooperative binding.

These experiments were repeated under non-inducing conditions. In marked contrast to the results obtained in galactose-containing medium, the ChIP signals representing native Gal4 occupancy of both the GAL1/10 and GAL3 promoters decayed rapidly, with a half-life of 5 min or less after the addition of β -oestradiol (Fig. 2b and Supplementary Fig. 3b). There was also a corresponding build-up of the signal due to formation of the Myc-G4-ER-VP16-promoter complex on each promoter with a similar time course (Fig. 2b). Dissociation of the Gal4-promoter complexes under non-inducing conditions was confirmed by measuring the rate of transcriptional induction after the addition of β -oestradiol. In glycerol- and lactic-acid-containing media, native Gal4 does not drive transcription owing to Gal80-mediated repression. However, as Myc-G4-ER-VP16 lacks the native Gal4 activation domain and therefore does not bind Gal80, it will act as a constitutive activator once bound to GAL promoters. As seen in Fig. 2c, the induction of transcription of the GAL1 gene in non-inducing media after addition of β -oestradiol to the cells was rapid and correlated well with the rate of binding detected by ChIP. This burst of transcription clearly represents specific association of the Gal4 DNA-binding domain (DBD)-containing competitor protein with the GAL1/10 promoter. When the same experiment was repeated with a competitor protein that contained the Put3, rather than the Gal4, DBD, no induction of GAL1 expression was detected (Fig. 2c). We conclude that the Gal4-promoter complexes are far more labile under non-inducing conditions than is the case when the activator is driving transcription. Indeed, the act of transcriptional activation seems to 'lock in' the Gal4-promoter complex.

In order to test the requirement for activated transcription to stabilize the Gal4-DNA interaction, the competition experiment was repeated in cells expressing a truncated Gal4 protein lacking the entire C-terminal activation domain (Gal4(1-841)) that is unable to bind to the transcription complex. Under inducing conditions, before the addition of β -oestradiol, Gal4(1-841) was present on the GAL1/10 promoter as determined by ChIP assay and little or no occupancy of the Myc-G4-ER-VP16 fusion protein was observed, as expected (Supplementary Fig. 5a). However, after addition of β -oestradiol, the competitor protein occupied the promoter rapidly (Supplementary Fig. 5a). These results were confirmed by measuring the precipitated DNA by quantitative polymerase chain reaction (Supplementary Fig. 5b). The production of GAL1 transcript was also consistent with the rapid replacement of Gal4(1-841) by the fusion protein. Little or no transcription was observed before β -oestradiol addition, but transcription was induced rapidly on steroid addition, reflecting replacement of the inactive Gal4(1-841) with the competitor protein, with its potent VP16 activation domain (AD) (Supplementary Fig. 5c). We conclude that the highly stable wild-type Gal4-promoter complexes observed under inducing conditions are a consequence of the activator driving transcription and are not the result of an indirect effect of changing the sugar source from glycerol/lactic acid to glucose.

The results presented above are difficult to reconcile with a mandatory coupling of ubiquitin/proteasome pathway (UPP)-mediated recycling of Gal4-promoter complexes with Gal4 activity, as has been suggested previously^{11,12}. Therefore, we tested directly the sensitivity of Gal4-mediated activation of GAL1 gene expression to the potent proteasome inhibitor MG132 in two different strains carrying mutations (*Dis1* (ref. 19) and *Apr5* (ref. 11)) that allow this compound to enter the cell. As shown in Fig. 3a, c, both the levels and kinetics of induction of GAL1 expression (shown normalized to ACT1 expression) were similar in the presence and absence of MG132 in both strains. Proteasome inhibition was confirmed by peptidase

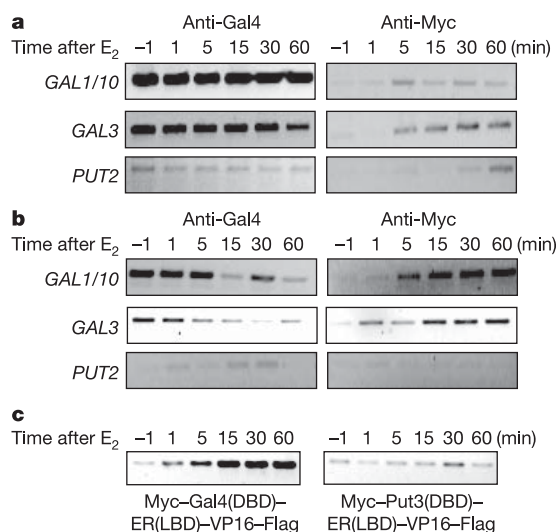


Figure 2 | Kinetic stability of Gal4-promoter complexes under inducing and non-inducing conditions. **a**, Competition ChIP assays were performed using GAL4 cells that overexpress Myc-G4-ER-VP16 and were grown in minimal medium with galactose as the carbon source (inducing conditions). β -Oestradiol (E₂) was added at $t = 0$. **b**, Same as **a**, but the cells were grown in glycerol and lactic acid (non-inducing conditions). **c**, Either Myc-G4-ER-VP16 or Myc-Put3(DBD)-ER(LBD)-VP16-Flag was expressed in GAL4 cells. The levels of GAL1 transcript were then determined by RT-PCR before and after the addition of β -oestradiol to yeast grown in glycerol/lactic acid (non-inducing) media.

assay and immunoblotting to detect the increase of polyubiquitinated proteins on treatment with MG132 (Supplementary Fig. 6). The same result was obtained when Gal4-induced α -galactosidase (the product of the *MEL1* gene) activity was measured in *ise1* cells (Fig. 3b), demonstrating that RNA processing steps downstream of transcription are also unaffected by the inhibitor.

These results do not support conclusions reached by a previous study¹¹. In a study focused primarily on the yeast transcription factor Gcn4, these workers presented a small amount of evidence that indicated that proteasome activity is essential for Gal4-mediated transcription. We do not understand the reason for the difference between our data for Gal4 and that reported previously, some of which was done in the same strain ($\Delta pdr5$)¹¹. The data presented here also do not support a model proposed by ref. 12, that UPP-mediated turnover of transcriptionally active Gal4 is essential for post-transcriptional events in *GAL* gene expression^{12,20}. We note that this model was based on indirect evidence and correlations, whereas the experiments reported here measure directly the lifetime of the Gal4–DNA complexes that drive *GAL* transcription. We suggest that the central experimental finding of ref. 12, which was that deletion of the gene encoding the E3 ubiquitin ligase Mdm30/Dsg1 crippled post-transcription steps in *GAL* gene expression, is due to the action of Mdm30/Dsg1 on some protein other than Gal4.

Studies of a few eukaryotic transcriptional activators have

suggested that transactivator–promoter complexes are highly dynamic in cells and that proteasome-mediated turnover is mechanistically coupled to activator function. The ‘competition ChIP’ assay that we developed (Fig. 1a) allows direct, time-resolved observation of exchange between native Gal4 and a competitor protein on native promoters. This avoids a limitation of the fluorescence recovery from photo-bleaching (FRAP) assay that has been used to monitor dynamics, which generally requires artificial promoters containing hundreds of activator-binding sites so as to form an observable ‘dot’²¹. The data obtained using this method (Fig. 2), along with the lack of an effect of MG132 on Gal4-mediated transcription (Fig. 3), show that proteolytic Gal4 turnover is not coupled mechanistically with continued transactivation of the *GAL* genes and therefore, this cannot be a general property of transcriptional activators. The Fig. 2 data are also inconsistent with rapid and reversible non-proteolytic dissociation of the activating Gal4–promoter complexes by chaperones, although this could be the case for the Gal4–Gal80 complex, which exchanges far more rapidly. We note that the accompanying paper, using completely different methodology, concluded that heat shock gene expression in *Drosophila* is driven by stable, long-lived HSF1–DNA complexes²². Furthermore, even for ER- α , where tight coupling between activation potency and UPP-mediated proteolysis has been reported, these events can be decoupled by mutation²³. Therefore, the body of data now available argues that some potent transactivators, such as Gal4 and HSF, function via long-lived complexes with promoters whereas others may be turned over more rapidly. Understanding the molecular mechanisms that regulate transcription factor–promoter lifetimes and why proteasome function is required for the activity of some activators but not others remains an important challenge.

METHODS

Competitive ChIP assays. Yeast cells expressing Myc–G4–ER–VP16 were grown in inducing or non-inducing media to an absorbance of 0.5–0.8 at 600 nm. One-hundred-microlitre aliquots were taken 1 min before and at the indicated times in the experiment after the addition of 1 μ M β -oestradiol and cross-linked in 1% formaldehyde for 15 min. Cross-linked cells were centrifuged for 5 min at 3,000g in a Sorvall RT7 centrifuge with an RTH-750 swing bucket rotor. Cells were washed with PBS and centrifuged as before. Cell pellets were frozen in liquid nitrogen and stored at -80°C until chromatin was prepared. Chromatin preparations and immunoprecipitations were performed as described²⁴. Chromatin was immunoprecipitated with anti-Gal4 C-terminal antibody bound to protein A agarose or anti-c-Myc-conjugated agarose. Chromatin was precipitated with goat anti-rabbit IgG and protein A agarose alone as controls. Other methods can be found in Supplementary Information.

Received 23 April; accepted 10 July 2006.

Published online 23 August 2006.

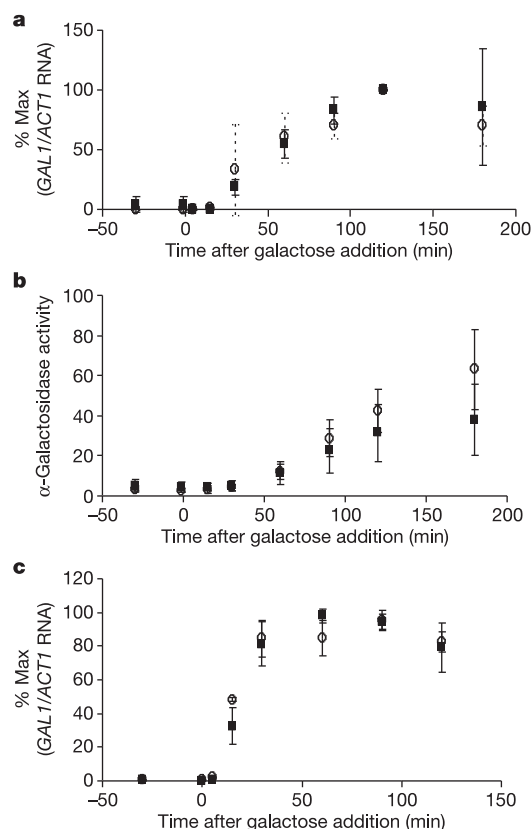


Figure 3 | Effect of MG132, an inhibitor of 26S proteasome-mediated proteolysis, on Gal4-mediated gene expression. MG132 in DMSO (filled squares), or DMSO alone (open circles), was added to *Δise1* cells, which are sensitive to proteasome inhibitors, grown in raffinose media (non-inducing) 30 min before the addition of galactose (induction). **a**, Levels of *GAL1* transcript (normalized to *ACT1* mRNA) as determined by quantitative RT–PCR at the times indicated after galactose addition. **b**, Same as **a**, except that α -galactosidase activity (the *MEL1* gene product) was measured from lysates made from cells at the time points indicated after galactose addition. Units of α -galactosidase activity are absolute Miller units. **c**, Same as **a**, but the $\Delta pdr5$ strain was used. Error bars represent s.e.m. in three independent experiments.

- McNally, J. G., Müller, W. G., Walker, D., Wolford, R. & Hager, G. L. The glucocorticoid receptor: rapid exchange with regulatory sites in living cells. *Science* **287**, 1262–1265 (2000).
- Freeman, B. C. & Yamamoto, K. R. Continuous recycling: a mechanism for modulatory signal transduction. *Trends Biochem. Sci.* **26**, 285–290 (2001).
- Elbi, C. *et al.* Molecular chaperones function as steroid receptor nuclear mobility factors. *Proc. Natl Acad. Sci. USA* **101**, 2876–2881 (2004).
- Fletcher, T. M. *et al.* ATP-dependent mobilization of the glucocorticoid receptor during chromatin remodeling. *Mol. Cell. Biol.* **22**, 3255–3263 (2002).
- Freeman, B. C., Felts, S. J., Toft, D. O. & Yamamoto, K. R. The p23 molecular chaperones act at a late step in intracellular receptor action to differentially affect ligand efficacies. *Genes Dev.* **14**, 422–434 (2000).
- Freeman, B. C. & Yamamoto, K. R. Disassembly of transcriptional regulatory complexes by molecular chaperones. *Science* **296**, 2232–2235 (2002).
- Métivier, R. *et al.* Estrogen receptor- α directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell* **115**, 751–763 (2003).
- Reid, G. *et al.* Cyclic, proteasome-mediated turnover of unliganded and liganded ER α on responsive promoters is an integral feature of estrogen signaling. *Mol. Cell* **11**, 695–707 (2003).
- Shang, Y., Hu, X., DiRenzo, J., Lazar, M. A. & Brown, M. Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. *Cell* **103**, 843–852 (2000).

10. Nawaz, Z. & O'Malley, B. W. Urban renewal in the nucleus: is protein turnover by proteasomes absolutely required for nuclear receptor-regulated transcription? *Mol. Endocrinol.* **18**, 493–499 (2004).
11. Lipford, J. R., Smith, G. T., Chi, Y. & Deshaies, R. J. A putative stimulatory role for activator turnover in gene expression. *Nature* **438**, 113–116 (2005).
12. Muratani, M., Kung, C., Shokat, K. M. & Tansey, W. P. The F box protein Dsg1/Mdm30 is a transcriptional coactivator that stimulates Gal4 turnover and cotranscriptional mRNA processing. *Cell* **120**, 887–899 (2005).
13. Fankhauser, C. P., Briand, P. A. & Picard, D. The hormone binding domain of the mineralocorticoid receptor can regulate heterologous activities in cis. *Biochem. Biophys. Res. Commun.* **200**, 195–201 (1994).
14. Picard, D. Posttranslational regulation of proteins by fusions to steroid-binding domains. *Methods Enzymol.* **327**, 385–401 (2000).
15. Louvion, J. F., Havaux-Copf, B. & Picard, D. Fusion of GAL4–VP16 to a steroid-binding domain provides a tool for gratuitous induction of galactose-responsive genes in yeast. *Gene* **131**, 129–134 (1993).
16. Wehrman, T. S., Casipit, C. L., Gewertz, N. M. & Blau, H. M. Enzymatic detection of protein translocation. *Nature Methods* **2**, 521–527 (2005).
17. Siddiqui, A. H. & Brandriss, M. C. The *Saccharomyces cerevisiae* PUT3 activator protein associates with proline-specific upstream activation sequences. *Mol. Cell. Biol.* **9**, 4706–4712 (1989).
18. Lohr, D., Venkov, P. & Zlatanova, J. Transcriptional regulation in the yeast GAL gene family: a complex genetic network. *FASEB J.* **9**, 777–787 (1995).
19. Lee, D. H. & Goldberg, A. L. Selective inhibitors of the proteasome-dependent and vacuolar pathways of protein degradation in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **271**, 27280–27284 (1996).
20. Arndt, K. & Winston, F. An unexpected role for ubiquitylation of a transcriptional activator. *Cell* **120**, 733–734 (2005).
21. Sprague, B. L. & McNally, J. G. FRAP analysis of binding: proper and fitting. *Trends Cell Biol.* **15**, 84–91 (2005).
22. Yao, J., Munson, K. M., Webb, W. W. & Lis, J. T. Dynamics of heat shock factor association with native gene loci in living cells. *Nature* doi:10.1038/nature05025 (this issue).
23. Valley, C. C. *et al.* Differential regulation of estrogen-inducible proteolysis and transcription by the estrogen receptor- α N terminus. *Mol. Cell. Biol.* **25**, 5417–5428 (2005).
24. Gonzalez, F., Delahodde, A., Kodadek, T. & Johnston, S. A. Recruitment of a 19S proteasome subcomplex to an activated promoter. *Science* **296**, 548–550 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This research was supported by the National Institutes of Health and the NHLBI Proteomics Initiative of the National Heart, Lung and Blood Institute, NIH. K.N. was supported by an NIH Cardiology Training Grant Fellowship. ER(LBD)-encoding plasmids were a gift from D. Picard.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.K. (Thomas.Kodadek@utsouthwestern.edu).

LETTERS

Nutrient regulates Tor1 nuclear localization and association with rDNA promoter

Hong Li^{1*}, Chi Kwan Tsang^{1*}, Marcus Watkins^{2*}, Paula G. Bertram² & X. F. Steven Zheng¹

TOR is the target of the immunosuppressant rapamycin and a key regulator of cell growth. It modulates diverse cellular processes in the cytoplasm and nucleus^{1–5}, including the expression of amino acid transporters, ribosomal RNAs and ribosomal proteins. Despite considerable recent progress, little is known about the spatial and temporal regulation of TOR signalling, particularly that leading into the nucleus. Here we show that Tor1 is dynamically distributed in the cytoplasm and nucleus in yeast. Tor1 nuclear localization is nutrient dependent and rapamycin sensitive: starvation or treatment with rapamycin causes Tor1 to exit from the nucleus. Tor1 nuclear localization is critical for 35S rRNA synthesis, but not for the expression of amino acid transporters and ribosomal protein genes. We show further that Tor1 is associated with 35S ribosomal DNA (rDNA) promoter chromatin in a rapamycin- and starvation-sensitive manner; this association is necessary for 35S rRNA synthesis and cell growth. These results indicate that the spatial regulation of TOR complex 1 (TORC1) might be involved in differential control of its target genes. TOR is known as a classic cytoplasmic kinase that mediates the cytoplasm-to-nucleus signalling by controlling the localization of transcription factors. Our data indicate that TOR might be more intimately involved in gene regulation than previously thought.

To study the spatial distribution of the rapamycin-sensitive TORC1 (refs 6, 7), we generated an antibody against the unique amino terminus of yeast Tor1. This antibody recognizes Tor1 by western blotting and indirect immunofluorescence, but not the closely related Tor2 (Supplementary Fig. 1a–c). In normally growing yeast cells, Tor1 is found in a punctate pattern in both the cytoplasm and nucleus, visibly more concentrated in the nucleus (Fig. 1a, b). This localization pattern is seen in two common budding yeast strains, W303 and S288C/FM391 (Supplementary Fig. 1b, c). The nuclear Tor1 is sensitive to rapamycin or starvation, which significantly decreases Tor1 nuclear staining (Fig. 1a, b). Under these conditions, cytoplasmic Tor1 distribution also changes from the punctate pattern to more tubular structures (Fig. 1a, b). To confirm Tor1 nuclear localization, we performed subcellular fractionation. TOR proteins have previously been shown to associate with intracellular membranes⁷, probably that of the endoplasmic reticulum and mitochondria^{8,9}. To separate the nuclei from the endoplasmic reticulum and mitochondria, we performed sucrose gradient ultracentrifugation¹⁰. Fraction 7 was enriched with nuclei yet largely free of endoplasmic reticulum and mitochondria contamination (Fig. 1c). Tor1 was found in fraction 7 from cells before, but not after, rapamycin treatment. Because rapamycin does not affect the overall level of Tor1 proteins (Tot, Fig. 1c), the immunofluorescence and subcellular fractionation results together indicate that Tor1 is normally localized in both the cytoplasm and the nucleus, and that rapamycin or nutrient starvation causes Tor1 to exit from the nucleus.

Protein nuclear transport is mediated by the interaction of nuclear localization sequence (NLS) and nuclear exporting sequence (NES) of the cargoes with importin and exportin (collectively called karyopherins), respectively¹¹. To identify Tor1 nuclear transporters, we have screened mutants of all 14 karyopherins encoded by the yeast genome¹². Two specific karyopherin mutants are found to affect Tor1 localization: Tor1 is enriched in the nucleus in the exportin *crm1-1^{ts}* mutant and its nuclear localization is resistant to rapamycin (Fig. 2a, and Supplementary Fig. 2a); in contrast, Tor1 is excluded from the

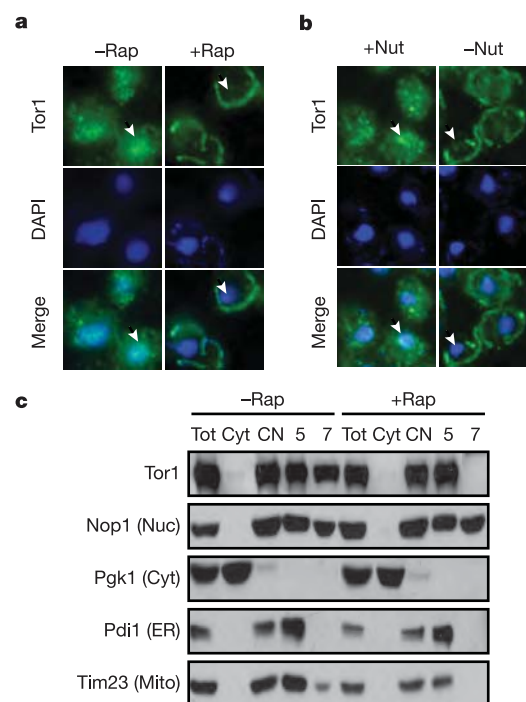


Figure 1 | Tor1 is dynamically distributed in the cytoplasm and nucleus in a rapamycin- and nutrient-sensitive manner. **a, b**, Tor1 nuclear localization is sensitive to rapamycin and to nutrient starvation. Yeast cells growing exponentially were treated without or with rapamycin (+Rap) (**a**) or shifted to YP medium (–Nut) (**b**) for 1 h, and were analysed for Tor1 localization by immunofluorescence. **c**, Tor1 localization is determined by subcellular fractionation. The marker proteins were Nop1, nucleus (Nuc); Pgk, cytosol (Cyt); Pdi1, endoplasmic reticulum (ER) and Tim 23, mitochondria (Mito). Tot, total cell lysate; Cyt, cytosol; CN, crude nuclei (also containing endoplasmic reticulum and mitochondria); 5, fraction 5 containing endoplasmic reticulum, mitochondria and nuclei; 7, fraction 7 containing pure nuclei.

¹Department of Pharmacology, Robert Wood Johnson Medical School, 675 Hoes Lane, Piscataway, New Jersey 08854, USA. ²Department of Medicine, Washington University School of Medicine, 660 South Euclid Avenue, St Louis, Missouri 63110, USA.

*These authors contributed equally to this work.

nucleus in the importin *srp1^{cs}* mutant even in the absence of rapamycin (Fig. 2b, and Supplementary Fig. 2b). By scanning the Tor1 sequence with the program PSORT, we identified a leucine-rich NES (¹⁰³¹LVPLTLTLFL) for Crm1, and two consensus NLSs (NLS1,

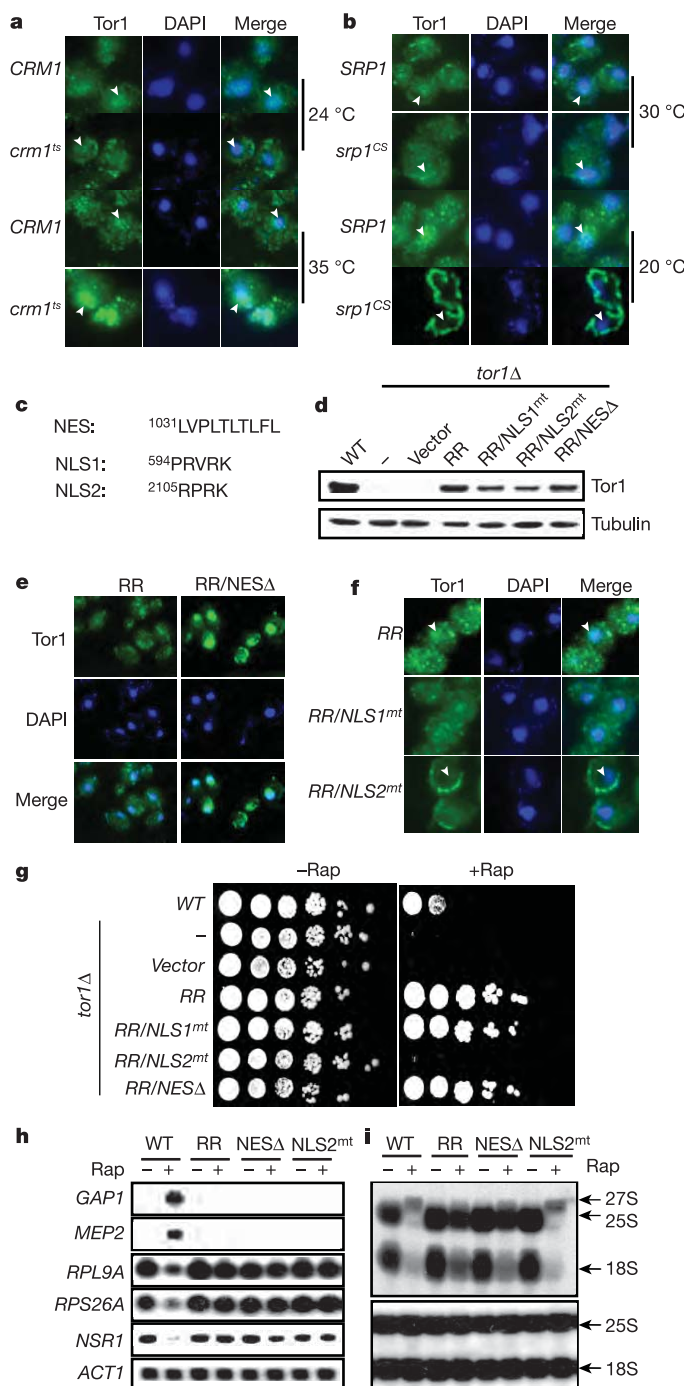


Figure 2 | Tor1 nuclear localization is critical for cell growth and 35S rRNA synthesis. **a, b**, *ts* alleles of *CRM1* and *SRP1* block Tor1 nuclear export and import, respectively. **c–f**, Tor1 NES and NLS. **c**, Tor1 carries several putative NES and NLS motifs. **d**, *TOR1-RR* alleles are expressed at similar levels. WT, wild type. **e**, Deletion of NES causes enrichment of Tor1 in the nucleus. **f**, Mutation of NLS2 prevents Tor1 nuclear localization. **g**, Tor1 nuclear localization is required for rapamycin-sensitive growth. **h**, Tor1 nuclear localization is not required for TORC1-dependent, Pol II-transcribed genes. **i**, Tor1 nuclear localization is required for 35S rRNA synthesis. 35S rRNA and processed products were pulse-labelled with [³H]methionine (top). Bottom: methylene blue staining of 25S and 18S rRNAs as a loading control.

⁵⁹⁴PRVRK; NLS2, ²¹⁰⁵RPRK) for Srp1 (Fig. 2c). To test whether these motifs are involved in Tor1 transport, we deleted the NES from a plasmid-borne *TOR1* rapamycin-resistant (*TOR1-RR*) allele. Tor1-RR contains the Ser 1972 → Ile mutation in the FKBP12–rapamycin-binding (FRB) domain that confers dominant rapamycin resistance¹³. As expected, Tor1-RR nuclear localization is not inhibited by rapamycin (Supplementary Fig. 1d). The *TOR1-RR* allele is useful for characterization of both Tor1 nuclear localization and functions (see below). Because NLS2 is near the kinase domain, we chose point mutations (⁵⁹⁴PR → AG and ²¹⁰⁷RK → IL) to disable the two NLSs, with the intention of minimizing the potentially deleterious effect on the kinase domain. Figure 2d shows that different Tor1-RR variants are expressed at similar levels. Tor1-RR is localized in both the cytoplasm and nucleus, but NES deletion causes Tor1-RR accumulation in the nucleus (Fig. 2e). In addition, Tor1(NESΔ) nuclear localization is insensitive to rapamycin (Supplementary Fig. 2c). Conversely, a mutation in NLS2, but not in NLS1, inhibits Tor1-RR nuclear localization (Fig. 2f, and Supplementary Fig. 2d). Taken together, these observations indicate that Crm1 and NES might mediate Tor1 nuclear export, and that Srp1 and NLS2 are responsible for Tor1 nuclear import. In addition, the *srp1^{cs}* and NLS2^{mt} mutations also change Tor1 localization from the punctate pattern to more tubular structures in the cytoplasm, indicating that Tor1 cytoplasmic distribution might be somehow tied to its nucleocytoplasmic export.

We next investigated the functional importance of Tor1 nuclear localization with the use of *TOR1-RR* alleles defective in nuclear transport. Rapamycin inactivates endogenous, wild-type *TOR*s, but not plasmid-borne *TOR1-RR* alleles, thereby allowing the specific examination of the ability of *TOR1-RR* alleles to regulate gene expression. Supplementary Fig. 3a illustrates the experimental strategy and potential outcomes. As shown previously¹⁴, rapamycin significantly inhibits the growth of the wild-type or *tor1Δ* strain, but not the wild-type or *tor1Δ* strain carrying *TOR1-RR* (Fig. 2g, and Supplementary Fig. 3b). *TOR1-RR(NESΔ)* and *TOR1-RR(NLS1^{mt})* but not *TOR1-RR(NLS2^{mt})* cells are rapamycin resistant, demonstrating that Tor1 nuclear localization is critical for Tor1-RR to confer rapamycin-resistant growth. To investigate the effect of Tor1 nuclear localization on TORC1-dependent genes, we analysed several TORC1-dependent RNA polymerase II (Pol II)-transcribed genes, including *GAP1*, *MEP2*, *RPL9A*, *RPS26A* and *NSR1*. Rapamycin is known to inhibit *RPL9A*, *RPS26A* and *NSR1* but derepress *GAP1* and *MEP2* (Fig. 2h)^{15–17}. Like *Tor1-RR*, *TOR1-RR(NESΔ)* and *TOR1-RR(NLS2^{mt})* also have the rapamycin-resistant phenotype for these genes, indicating that Tor1 nuclear localization is not absolutely required for the regulation of Pol II-dependent genes. Because Tor1 kinase activity is essential for Tor1 to inhibit *GAP1* (ref. 15), the fact that *TOR1-RR(NESΔ)* and *TOR1-RR(NLS2^{mt})* are still capable of repressing *GAP1* expression indicates that *NESΔ* and *NLS2^{mt}* mutations do not interfere with Tor1 kinase activity.

TOR is also known to regulate Pol I-dependent 35S ribosomal DNA transcription^{18,19}, a key event for ribosome biogenesis and cell growth^{20,21}. We therefore investigated the role of Tor1 nuclear localization in 35S rRNA synthesis and processing by radiolabelling 35S rRNA with [³H]methionine²². In the wild-type strain in the absence of rapamycin, the predominantly labelled species are 25S and 18S rRNAs, because newly synthesized 35S rRNA is rapidly processed into the mature rRNAs²² (Fig. 2i). Rapamycin strongly inhibits 35S rRNA synthesis as demonstrated by the decrease in ³H-labelled 25S and 18S rRNAs. There is also a slight increase in ³H-labelled 35S and 27S rRNAs as a result of inhibition of 35S rRNA processing (Fig. 2i; see Fig. 4e, below, for 35S rRNA). *TOR1-RR* strongly resists rapamycin, as indicated by the relatively normal amount of ³H-labelled 25S and 18S rRNAs and by the absence of ³H-labelled 35S and 27S rRNAs. However, 35S rRNA synthesis and processing are strongly inhibited by rapamycin in the *TOR1-RR(NLS2^{mt})*, but not the *TOR1-RR(NESΔ)*, strain. Together, these data show that Tor1 nuclear localization is important for 35S rRNA synthesis.

The essential role of Tor1 nuclear localization in 35S rRNA synthesis indicates that Tor1 might be more directly involved in rDNA regulation. We therefore investigated the interaction of Tor1 with rDNA promoter by chromatin immunoprecipitation (ChIP). After protein–DNA crosslinking and anti-Tor1 immunoprecipitation, Tor1-associated DNA was analysed by polymerase chain reaction (Fig. 3a). In *tor1Δ* strain or with a control antibody, there is little or no detectable DNA precipitation, demonstrating the specificity of the ChIP assay (Fig. 3b). In exponential cells, Tor1 is detected on the 35S rDNA promoter but not on the 5.8, 18 and 25S coding regions. Additionally, Tor1 is not detected at *CUP1*, *ACT1* (Fig. 3b) or several ribosomal protein loci (*RPS6A*, *RPL9A* and *RPS26A*; Fig. 3d). Rapamycin causes the dissociation of Tor1, but not Tor1-RR, from the 35S rDNA promoter (Fig. 3b, c). We further mapped the intergenic region for the precise Tor1-binding site(s) (Fig. 3a). Tor1 binding occurs in regions containing or adjacent to the 35S rDNA promoter in a rapamycin-sensitive manner (Fig. 3e, f), with

peak Tor1-binding activity at the 35S rDNA promoter. In addition, 5S rDNA also shows Tor1 binding. Tor1 might therefore also interact with 5S rDNA promoter that is embedded in 5S rDNA transcribed region. Starvation decreases the association of both Tor1 and Tor1-RR with rDNA promoter (Fig. 3g). Consistent with their subcellular localization is our finding of Tor1-RR(NESΔ), but not Tor1-RR(NLS2^{mt}), at the rDNA promoter site (Fig. 3g). More Tor1-RR(NESΔ) than Tor1 or Tor1-RR seems to be present at the rDNA promoter (Fig. 3g), and nutrient starvation causes only a slight decrease in Tor1-RR(NESΔ) from the 35S rDNA promoter, indicating that the interaction of Tor1 with rDNA promoter might be regulated by both nuclear import and an import-independent mechanism. Kog1, a component of TORC1, but not Avo2/3, components of TORC2, is associated with the 35S rDNA promoter in a rapamycin-sensitive manner (Fig. 3h). It seems that TORC1, but not TORC2, interacts with the 35S rDNA promoter.

Analysis of Tor1 sequence revealed three classic DNA-binding motifs: two leucine zippers (LZ1, residues 277–291, and LZ2, residues 311–332) and a conserved helix–turn–helix (HTH, residues 815–837) (Fig. 4a). To investigate the importance of these motifs, we individually deleted the two leucine zippers and HTH from *TOR1-RR*. These *TOR1-RR* alleles are expressed at similar levels in yeast (Fig. 4b). By ChIP assay, we found that the deletion of HTH, but not

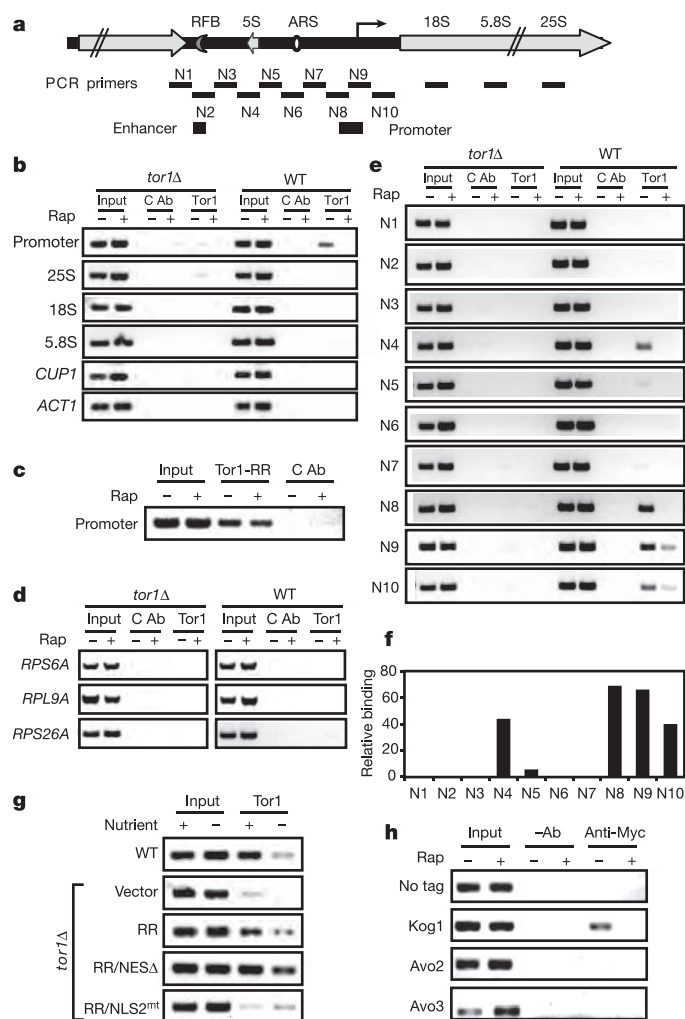


Figure 3 | Tor1 binds to the 35S rDNA promoter in a rapamycin-sensitive manner. **a**, The structure of the yeast rDNA repeat and the PCR primer sets. **b**, Tor1 binds to the 35S rDNA promoter in a rapamycin-sensitive manner as determined by ChIP. C Ab, control antibody; WT, wild type. **c**, Tor1-RR binding to the 35S rDNA promoter is resistant to rapamycin. **d**, Tor1 is not detectable at the promoter of ribosomal protein genes. **e**, Mapping of Tor1 binding site(s) in the intergenic rDNA regions. **f**, Quantification of Tor1 binding to different rDNA intergenic regions. **g**, The effect of nutrient starvation on the promoter binding of different Tor1 proteins. **h**, TORC1, but not TORC2, interacts with the 35S rDNA promoter. The ability of Kog1-MYC9, Avo2-MYC9 or Avo3-MYC9 to bind to the 35S rDNA promoter was assayed by ChIP with a Myc-specific antibody. –Ab, no antibody control.

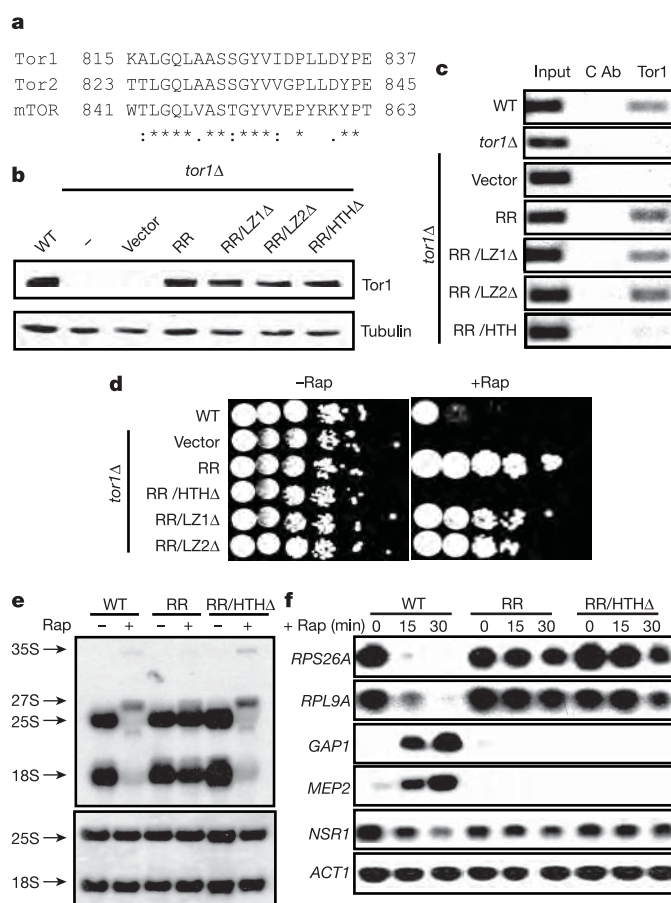


Figure 4 | Tor1 binding to the 35S rDNA promoter is important for cell growth and 35S rRNA synthesis. **a–c**, HTH, but not LZ1 and LZ2, is required for Tor1 binding to rDNA promoter. **a**, HTH is conserved among *S. cerevisiae* Tor1 and Tor2, and mTOR. **b**, Expression of different *TOR1-RR* alleles in the *tor1Δ* strain. **c**, Deletion of HTH abrogates the ability of Tor1 to bind to the 35S rDNA promoter. C Ab, control antibody; WT, wild type. **d**, HTH is required for rapamycin-sensitive cell growth. **e**, HTH is critical for 35S rRNA synthesis. 35S rRNA synthesis and processing are analysed as in Fig. 2i. **f**, HTH is not required for the expression of TORC1-dependent, RNA polymerase II-transcribed genes. *ACT1* was used as a loading control.

that of LZ1 and LZ2, abrogates the ability of Tor1-RR to bind to rDNA promoter, indicating that HTH, but not LZ1 or LZ2, is required for binding to the 35S rDNA promoter (Fig. 4c). *TOR1-RR*, *TOR1-RR(LZ1Δ)* and *TOR1-RR(LZ2Δ)*, but not *TOR1-RR(HTHΔ)*, confer resistance to rapamycin (Fig. 4d), indicating that interaction of Tor1 with the 35S rDNA promoter might be important for growth. Moreover, deletion of HTH selectively abolishes the ability of Tor1-RR to promote 35S rRNA synthesis but not the expression of Pol II genes in the presence of rapamycin (Fig. 4e, f).

Taken together, our results indicate the following possible model for Tor1 dynamic distribution in the cytoplasm and nucleus in response to nutrient availability (Supplementary Fig. 4). Tor1 normally binds to the 35S rDNA promoter, directly through the HTH motif or indirectly through other DNA-binding protein(s); this binding is important for the regulation of 35S rRNA synthesis, possibly by modulating phosphorylation of the Pol I machinery²³ and/or condensation of the rDNA repeats²⁴. During starvation or treatment with rapamycin, Tor1 is dissociated from the rDNA promoter and exits from the nucleus. Although nuclear localization is apparently important for the association of Tor1 with rDNA promoter, there seems to be an import-independent mechanism that regulates the DNA-binding activity of Tor1. An intact Tor1 kinase domain is important for the localization of Tor1 to the nucleus because kinase-inactive Tor1-RR(D2294E) is found primarily in the cytoplasm (Supplementary Fig. 5). In contrast, DNA binding is not absolutely required for the retention of Tor1 in the nucleus (Supplementary Fig. 6). In contrast to 35S rRNA synthesis, the cytoplasmic form of Tor1 is apparently sufficient to regulate TORC1-dependent, Pol II-transcribed genes, including Gln3-dependent genes^{15,25,26} and Iff1:Fhl1-dependent ribosomal protein genes^{27–29}. These results indicate that TOR might use distinct mechanisms in the control of diverse gene expression in response to environmental changes.

METHODS

Indirect immunofluorescence microscopy. Yeast immunofluorescence studies were performed as described³⁰. Anti-Tor1 antibody was used at a dilution of 1:400. The antibody–antigen complexes were detected with Alexa Fluor 488-conjugated secondary antibody. DNA was stained for 15 min with 50 ng ml^{−1} 4',6-diamidino-2-phenylindole (DAPI) in antifade mounting medium (Vector Laboratories). Fluorescent signals were analysed with an Olympus fluorescence microscope equipped with a digital camera.

Other methods. Details for other methods can be found in Supplementary Information.

Received 25 January; accepted 28 June 2006.

Published online 9 August 2006.

1. Abraham, R. TOR signaling: an odyssey from cellular stress to the cell growth machinery. *Curr. Biol.* **15**, R139–R141 (2005).
2. Harris, T. E. & Lawrence, J. C. Jr. TOR signaling. *Sci. STKE* **2003**, re15 (2003).
3. Shamji, A., Nghiem, P. & Schreiber, S. Integration of growth factor and nutrient signaling: implications for cancer biology. *Mol. Cell* **12**, 271–280 (2003).
4. Schmelzle, T. & Hall, M. N. TOR, a central controller of cell growth. *Cell* **103**, 253–262 (2000).
5. Schaner, L., Duvel, K. & Broach, J. Sense and sensibility: nutritional response and signal integration in yeast. *Curr. Opin. Microbiol.* **7**, 624–630 (2004).
6. Loewith, R. et al. Two TOR complexes, only one of which is rapamycin sensitive, have distinct roles in cell growth control. *Mol. Cell* **10**, 457–468 (2002).
7. Wedaman, K. P. et al. Tor kinases are in distinct membrane-associated protein complexes in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **14**, 1204–1220 (2003).
8. Drenan, R. M., Liu, X., Bertram, P. G. & Zheng, X. F. S. FKBP12-rapamycin-associated protein or mammalian target of rapamycin (FRAP/mTOR) localization in the endoplasmic reticulum and the Golgi apparatus. *J. Biol. Chem.* **279**, 772–778 (2004).
9. Desai, B. N., Myers, B. R. & Schreiber, S. L. FKBP12-rapamycin-associated protein associates with mitochondria and senses osmotic stress via

- mitochondrial dysfunction. *Proc. Natl Acad. Sci. USA* **99**, 4319–4324 (2002).
10. Rout, M. & Kilmartin, J. in *Cell Biology: A Laboratory Handbook* (ed. Celis, J. E.), 2nd edn 120–128 (Academic, San Diego, 1998).
11. Pemberton, L. F., Blobel, G. & Rosenblum, J. S. Transport routes through the nuclear pore complex. *Curr. Opin. Cell Biol.* **10**, 392–399 (1998).
12. Carvalho, J., Bertram, P. G., Wenthe, S. R. & Zheng, X. F. S. Phosphorylation Regulates the Interaction between Gln3p and the Nuclear Import Factor Srp1p. *J. Biol. Chem.* **276**, 25359–25365 (2001).
13. Zheng, X. F., Florentino, D., Chen, J., Crabtree, G. R. & Schreiber, S. L. TOR kinase domains are required for two distinct functions, only one of which is inhibited by rapamycin. *Cell* **82**, 121–130 (1995).
14. Chan, T. F., Carvalho, J., Riles, L. & Zheng, X. F. S. A chemical genomics approach toward understanding the global functions of TOR. *Proc. Natl Acad. Sci. USA* **97**, 13227–13232 (2000).
15. Bertram, P. G. et al. Tripartite regulation of Gln3p by TOR, Ure2p and phosphatases. *J. Biol. Chem.* **275**, 35727–35733 (2000).
16. Cardenas, M., Cutler, N., Lorenz, M., Di Como, C. & Heitman, J. The TOR signaling cascade regulates gene expression in response to nutrients. *Genes Dev.* **13**, 3271–3279 (1999).
17. Jorgensen, P. et al. A dynamic transcriptional network communicates growth potential to ribosome synthesis and critical cell size. *Genes Dev.* **18**, 2491–2505 (2004).
18. Zaragoza, D., Ghavidel, A., Heitman, J. & Schultz, M. C. Rapamycin induces the G0 program of transcriptional repression in yeast by interfering with the TOR signaling pathway. *Mol. Cell. Biol.* **18**, 4463–4470 (1998).
19. Powers, T. & Walter, P. Regulation of ribosome biogenesis by the rapamycin-sensitive TOR-signaling pathway in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **10**, 987–1000 (1999).
20. Moss, T. & Stefanovsky, V. Y. At the center of eukaryotic life. *Cell* **109**, 545–548 (2002).
21. Warner, J. R. The economics of ribosome biosynthesis in yeast. *Trends Biochem. Sci.* **24**, 437–440 (1999).
22. Udem, S. & Warner, J. Ribosomal RNA synthesis in *Saccharomyces cerevisiae*. *J. Mol. Biol.* **65**, 227–242 (1972).
23. Claypool, J. A. et al. Tor pathway regulates Rrn3p-dependent recruitment of yeast RNA polymerase I to the promoter but does not participate in alteration of the number of active genes. *Mol. Biol. Cell* **15**, 946–956 (2004).
24. Tsang, C. K., Bertram, P. G., Ai, W., Drenan, R. & Zheng, X. F. S. Chromatin-mediated regulation of nucleolar structure and RNA Pol I localization by TOR. *EMBO J.* **22**, 6045–6056 (2003).
25. Beck, T. & Hall, M. N. The TOR signaling pathway controls nuclear localization of nutrient-regulated transcriptional factors. *Nature* **402**, 689–692 (1999).
26. Duvel, K., Santhanam, A., Garrett, S., Schaner, L. & Broach, J. Multiple roles of Tap42 in mediating rapamycin-induced transcriptional changes in yeast. *Mol. Cell* **11**, 1467–1478 (2003).
27. Martin, D., Souillard, A. & Hall, M. TOR regulates ribosomal protein gene expression via PKA and the Forkhead transcription factor FHL1. *Cell* **119**, 969–979 (2004).
28. Rudra, D., Zhao, Y. & Warner, J. R. Central role of Iff1p–Fhl1p interaction in the synthesis of yeast ribosomal proteins. *EMBO J.* **24**, 533–542 (2005).
29. Schawaldner, S. et al. Growth-regulated recruitment of the essential yeast ribosomal protein gene activator Iff1. *Nature* **432**, 1058–1061 (2004).
30. Bertram, P. G. et al. Convergence of TOR-nitrogen and Snf1-glucose signaling pathways onto Gln3. *Mol. Cell. Biol.* **22**, 1246–1252 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. H. Cho for generating the KOG1-MYC9 strain, M. Rout and S. Wenthe for advice on the purification of yeast nuclei, and H. Sesaki and R. Jensen for Tim23 antibody. This work was supported by an NIH grant to X.F.S.Z.

Author Contributions The immunofluorescence experiments were performed by C.K.T. and M.W. Mutagenesis and rapamycin sensitivity assays were performed by H.L. and M.W. Subcellular fractionation, northern blotting and 35S rRNA assays were performed by H.L. ChIP assays were performed by C.K.T. and H.L. Tor1 antibody was generated and characterized by P.G.B. The data were analyzed and interpreted by H.L., C.K.T., M.W., P.G.B. and X.F.S.Z. The data images for the figures were prepared by H.L., C.K.T., M.W. and P.G.B. The manuscript was written by H.L., C.K.T. and X.F.S.Z.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to X.F.S.Z. (zhengst@umdnj.edu).

LETTERS

In situ structure of the complete *Treponema primitia* flagellar motor

Gavin E. Murphy¹, Jared R. Leadbetter² & Grant J. Jensen¹

The bacterial flagellar motor is an amazing nanomachine: built from approximately 25 different proteins, it uses an electrochemical ion gradient to drive rotation at speeds of up to 300 Hz (refs 1, 2). The flagellar motor consists of a fixed, membrane-embedded, torque-generating stator and a typically bidirectional, spinning rotor that changes direction in response to chemotactic signals. Most structural analyses so far have targeted the purified rotor^{3,4}, and hence little is known about the stator and its interactions. Here we show, using electron cryotomography of whole cells, the *in situ* structure of the complete flagellar motor from the spirochaete *Treponema primitia* at 7 nm resolution. Twenty individual motor particles were computationally extracted from the reconstructions, aligned and then averaged. The stator assembly, revealed for the first time, possessed 16-fold symmetry and was connected directly to the rotor, C ring and a novel P-ring-like structure. The unusually large size of the motor suggested mechanisms for increasing torque and supported models wherein critical interactions occur atop the C ring, where our data suggest that both the carboxy-terminal and middle domains of FlgG are found.

The bacterial flagellar motor excites considerable interest because of the ordered expression of its genes, its regulated self-assembly, the complex interactions of its many proteins, and its startling mechanical abilities. Stator proteins MotA and MotB form a ring of 'studs' within and above the inner membrane that couple the passage of protons across the membrane to the generation of torque^{1,2}. Above the membrane, MotB has a peptidoglycan-binding domain that presumably holds the stator in place by binding to the globally cross-linked peptidoglycan layer^{1,2}. Below the membrane, the cytoplasmic loops of MotA are believed to spin a wheel of FlgG molecules, which—like radial spokes—extend roughly parallel to the membrane from the rotor in the middle to just below MotA on the periphery¹. Proteinaceous P and L rings serve as bearings to facilitate the rotation of the rod within the peptidoglycan and outer membranes, respectively^{1,2}. Inside the cell and below FlgG lies the C ring, which regulates the direction of rotation in response to the chemotactic system^{1,2}.

Flagellar basal bodies containing the rotor, rod and sometimes the C ring have been purified and reconstructed by electron-cryomicroscopy-based single-particle analysis^{3,5,6}. The *Salmonella* rotor possessed 26-fold symmetry⁷, whereas the *Salmonella* C ring possessed a mean symmetry of 34 (ref. 8). Because the stators do not co-purify with the rotor, however, little is known about their structure and interactions with the rest of the motor. Patterns of stator studs have been seen in two-dimensional, freeze-etch images, but the interpretation of these images is difficult and the number of studs has been reported as either 12 or 16, depending on the species^{9–12}. Two-dimensional electron cryomicroscopy images of purified PomA–PomB complexes (homologues of MotA and MotB) from *Vibrio alginolyticus* have revealed a ~70-Å-long, thin extension above the membrane¹³.

Here we report the complete structure of the flagellar motor, including the stators, obtained by electron cryotomography. Fifteen

Treponema primitia cells frozen within thin layers of vitreous ice were imaged (Fig. 1a and Methods). *T. primitia* was chosen for its narrow diameter and interesting periplasmic flagella that emerge from each pole. Twenty motor particles were computationally extracted from the reconstructions, mutually aligned and averaged (Fig. 1b–e). In both the individual maps (Fig. 1d) and their average (Fig. 1e), the stator studs were clearly 16-fold symmetric around the rod. We checked for symmetry computationally in the other components, including the P collar (the density above the stator, as explained below), the rotor, the connections between the stators and C ring, and the C ring itself (Supplementary Fig. 1a, b). Presumably because

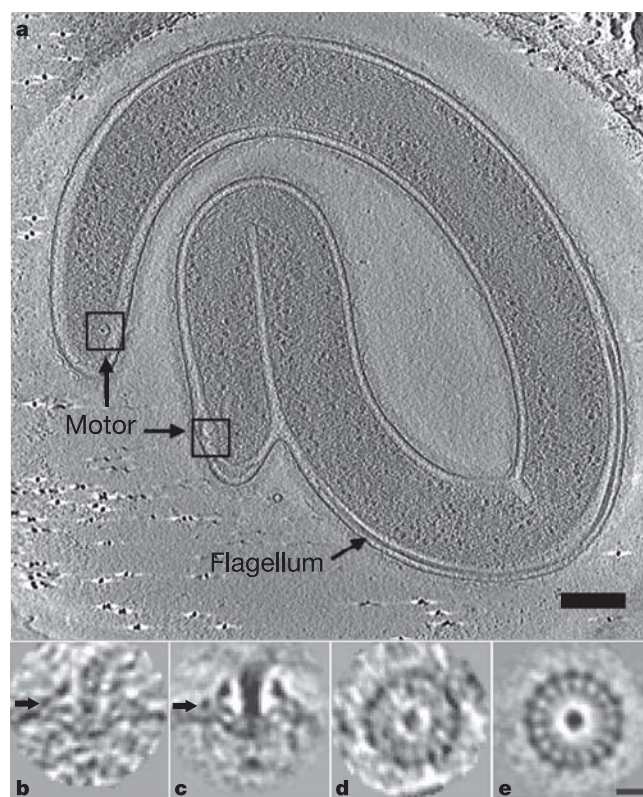


Figure 1 | Electron cryotomography of *T. primitia* and its periplasmic flagellar motor. **a**, A 2-nm-thick central section through a tomogram of an entire *Treponema* cell. A flagellar motor is located near each cell tip and the flagella rotate in the periplasm. Scale bar, 200 nm. **b**, Axial slice through the centre of one extracted motor particle. **c**, Axial slice through the average of twenty motor particles. **d**, Radial slice through the stator region of the same particle shown in **b** taken at the height indicated by the arrow in **b**. **e**, Radial slice through the average motor, taken at the height indicated by the arrow in **c**. Scale bar, 20 nm (for panels **b**–**e**).

¹Division of Biology and ²Division of Environmental Science and Engineering, California Institute of Technology, Pasadena, California 91125, USA.

of the limited resolution, only the symmetry of the stators and their connections to the C ring was apparent. This symmetry was therefore imposed on the entire motor, effectively smoothing the other components (Supplementary Fig. 1c, d).

For the first time, the three-dimensional structure of the stators was revealed in their natural position in contact with the membrane and other motor components (Fig. 2 and Supplementary Movie). The 16 stator studs (two of which are identified by asterisks in Fig. 2c) were 8 nm wide, which is similar to the reported values of 5–7 nm seen by other means^{9,10,12}. Surprisingly, the volume of each stud above the membrane was ~20-times larger than that expected for two MotB peptidoglycan-binding domains and much thicker than PomA–PomB resuspended in liposomes¹³. The identity of the rest of the stator density is unclear. The studs were spaced 7 nm apart, which is sufficient to accommodate hypothetical models for the 18 transmembrane helices in a (MotA)₄(MotB)₂ torque-generating unit¹⁴. The average stud was not vertical—instead, it leaned (dotted line in Fig. 2a) such that its distal end was positioned clockwise relative to the proximal end as viewed from inside the cell. There were thin, bridging densities connecting the stud heads around the ring (arrowhead in Fig. 2c).

Four bridging densities (numbered 1–4 in Fig. 2b) were seen connecting the stator (“S” in Fig. 2b) to other components of the motor (see Supplementary Fig. 2 for contour, variance and statistical significance maps). Bridging density 1 connected the stator to the C ring (“C” in Fig. 2b). It is thought that a series of charged residues in a cytoplasmic loop of MotA interacts here with complementary charges in the C-terminal domain of FliG¹. The stator–C-ring connections were also rotated with respect to the periplasmic studs, suggesting that they might perhaps be the terminus of a ~24-nm-long, straight component that extended from the peptidoglycan layer all the way through the membrane to the C ring (again, dotted line in Fig. 2a). Bridging densities 2 and 3 were finger-like extensions connecting the stator directly to the rotor.

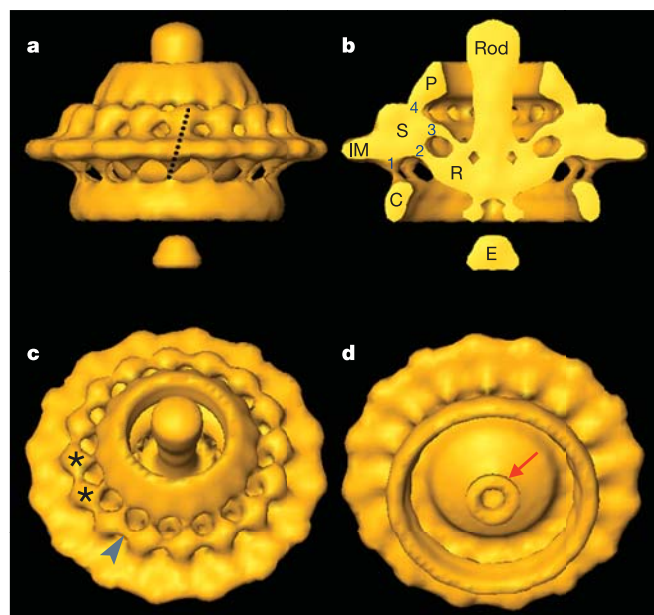


Figure 2 | Isosurface of the symmetrized average flagellar motor. **a**, Side view. The dotted line indicates the tilt of the stators with respect to the membrane. **b**, Cutaway view of the motor in the same orientation. The stators (S) are embedded in the inner membrane (IM) and directly contact the C ring (C; via bridging density 1), the rotor (R; via bridging densities 2 and 3) and the P collar (P; via bridging density 4). A rod extends from the rotor, and a bundle (E) lies under the rotor's pore. **c**, Oblique view of the motor's top from within the periplasm. Note the arrangement of the stator studs (asterisks) and their linkages (arrowhead). **d**, Oblique view of the motor's bottom from within the cytoplasm, with the bundle removed to reveal the pore ring (arrow).

Bridging density 4 linked the stator to a contrast-rich ring of density (“P” in Fig. 2b) encircling the rod above the rotor, reminiscent of the P- and L-ring bushings in *Salmonella* and *Escherichia coli*. Sequenced *Treponema* flagellar proteins are similar to the better-known *Salmonella* and *E. coli* versions (see Supplementary Figs 3–5), except that the genes for the P- and L-ring proteins FlgI and FlgH are missing¹⁵. The absence of an L ring is understandable because the periplasmic flagella of *Treponema* never exit the outer membrane, and, not surprisingly, isolated *Treponema* basal bodies lack any ring structures¹⁶. In our *in situ* reconstructions, however, an additional ring was seen just above the stators at the level of the peptidoglycan layer, but ~8 nm away (surface to surface) from the rod itself. We therefore refer to it as the ‘P collar’ to reflect its position and loose fit around the rod, though the gene responsible for this density is unknown. This structure may serve to limit the tilt of the flagellar hook and may also further stabilize the stators.

The rotor itself (“R” in Fig. 2b) was bowl-shaped. Unlike previous work on the isolated basal body, in which the membrane location was not certain^{3,7}, here the bulk of the rotor was seen to lie just beneath the inner membrane (“IM” in Fig. 2b) submerged within the C ring. At the bottom of the rotor, there was a small ring (arrow in Fig. 2d)

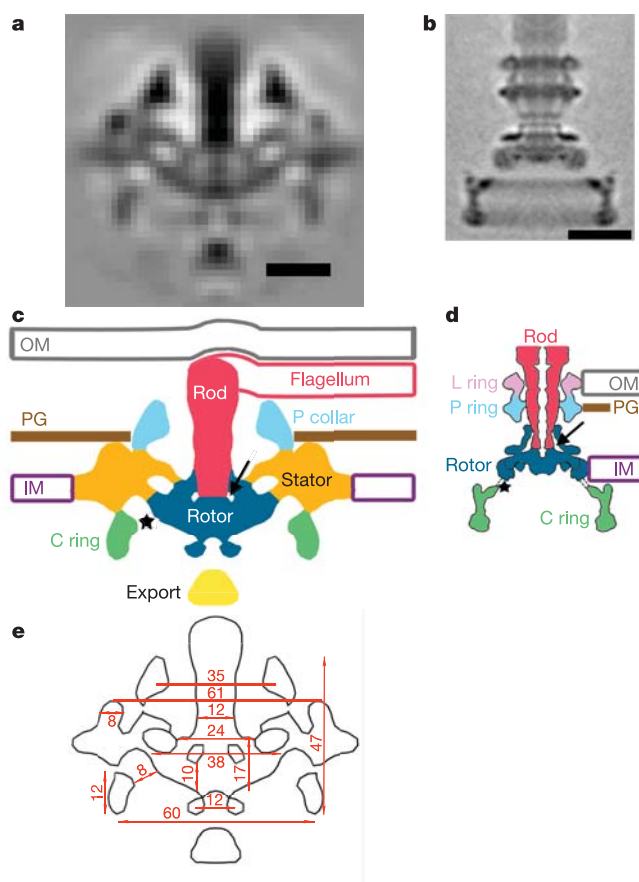


Figure 3 | The *Treponema* motor and its comparison with the *Salmonella* basal body. All objects are at the same scale. **a**, Axial slice through the *Treponema* flagellar motor. Scale bar, 20 nm. **b**, Averaged image of the *Salmonella* basal body. Scale bar, 20 nm. Modified from ref. 4. **c**, Cartoon interpretation of the *Treponema* motor with its components labelled. The arrows here and in **d** point to the gap between the rotor and rod, whereas the stars indicate the gap between the C ring and rotor. The location of the peptidoglycan layer (PG) is conjectured. The outlined objects' locations are approximate. The flagellum actually bends more gradually over the P collar in presumably random directions but appears straight when averaged. The inner membrane (IM) and outer membrane (OM) both bulge more widely around the motor *in situ* than is pictured. **d**, Cartoon of the *Salmonella* basal body for comparison. Modified from ref. 7. **e**, Measurements (in nm) of various motor features.

that formed a funnel-like pore, which may perhaps be the insertion apparatus through which flagellin monomers are exported. Another contrast-rich density with low variance (see Supplementary Fig. 2) was visible 4 nm below the pore and, because of its proximity, has been labelled an export bundle ("E" in Fig. 2b).

In comparison with isolated *Salmonella* basal bodies, which have been reconstructed to higher resolution by single-particle analysis^{3,5}, the *Treponema* stator ring, C ring and rotor are all much larger (Fig. 3). The rotor is also located lower within the C ring and appears bowl-shaped rather than disk-like. Notably, by stereo photogrammetry the *Salmonella* rotor also appeared bowl-shaped *in situ*¹⁷, so the shape may depend on conditions lost during purification. The *Caulobacter* rotor has also appeared bowl-shaped in some reconstructions⁶. Other structural details are remarkably conserved, such as the small gap between rotor and rod (arrows in Fig. 3c, d).

These differences have important implications for current models of the functional and architectural relationships of the components. Whereas the *Salmonella* motor spins just the flagellum, because *Treponema* flagella are periplasmic, it is thought that they cause the whole cell to gyrate¹⁸. Thus, each rotation may be much slower and require greater torque. The unusually large stud ring, C ring and rotor in *Treponema* may serve to increase torque by increasing the length of the effective lever arm through which each stator stud acts. These larger rings may also accommodate more stator studs and FliG molecules around the ring, in effect 'gearing down' the *Treponema* motor so that the passage of each proton across the membrane produces a smaller angular rotation.

FliG is thought to have three domains: a C-terminal domain directly underneath the stator that forms the top of the C ring, a middle domain whose location is uncertain, and an amino-terminal region bound to the rotor^{1,2,4,19,20}. In our reconstructions, the stator–C-ring connection appears on the outside edge of the C ring, and the distance between the C ring and the rotor is too large to be spanned by the 2-nm-long α -helix connecting the C-terminal and middle domains of FliG (see Supplementary Information and starred gap in Fig. 3). The simplest interpretation of these results is that the C-terminal domain forms the outside edge of the C ring, both the C-terminal and middle domains of FliG lie atop the C ring, and a portion of the N terminus acts as an extended tether spanning the gap to the rotor, as argued elsewhere^{19,20}. It is interesting to note that although the diameters of the stud and C rings in *Treponema* are unusually large, nevertheless they still match each other, so that the studs appear directly above the C ring. The available data suggest that the same relationship holds in other, smaller motors as well^{3,11}, supporting the idea that this juxtaposition is important and that key functional interactions do indeed occur at this interface.

METHODS

Additional details of the methods used in this study can be found in Supplementary Information.

Exponential phase cultures of *T. primitia* strain ZAS-2 (ref. 21) were plunge-frozen with gold fiducial markers across electron microscope grids in liquid ethane. *T. primitia* is an obligate anaerobe but can tolerate atmospheric conditions for about 20 min, so grids were frozen quickly in small batches. Single-axis tilt series were acquired automatically on a 300-keV FEI Polara field emission gun transmission electron microscope. The 20 averaged motors were taken from tilt series with underfocuses between 10 and 18 μm (first contrast transfer function (CTF) zero between 4.5 and 6.0 nm resolution). Tomograms were low-pass filtered at the resolution of the first CTF zero and binned once (1.96 nm pixel⁻¹). No other CTF corrections were performed.

Twenty motor particles were computationally extracted from the tomograms and aligned to an arbitrarily chosen reference particle. The aligned motors were then averaged and rotated so that the rod axis corresponded to the z-axis. To detect the symmetry of the components, annular masks were generated for the five different motor regions (shown in Supplementary Fig. 1b), and then applied separately to all twenty individual particles. The only detectable symmetry in the rotational power spectrum (16-fold) was found in the two stator regions (periplasmic and cytoplasmic), so this was applied to the entire averaged motor. This initial symmetrized average was used as a reference for a further

alignment, and the process was iterated a total of five times. To test potential reference bias, two alternative particles were used as a reference, and the resulting average was essentially identical. A resolution of 7 nm was estimated by separately averaging and symmetrizing two halves of the data set and correlating them using Fourier shell correlation with a threshold of 0.5. The isosurface was contoured at a level that showed the connections between stator and C ring.

Received 7 April; accepted 21 June 2006.

Published online 2 August 2006.

- Kojima, S. & Blair, D. F. The bacterial flagellar motor: structure and function of a complex molecular machine. *Int. Rev. Cytol.* **233**, 93–134 (2004).
- Berg, H. C. The rotary motor of bacterial flagella. *Annu. Rev. Biochem.* **72**, 19–54 (2003).
- Francis, N. R., Sosinsky, G. E., Thomas, D. & DeRosier, D. J. Isolation, characterization and structure of bacterial flagellar motors containing the switch complex. *J. Mol. Biol.* **235**, 1261–1270 (1994).
- Thomas, D., Morgan, D. G. & DeRosier, D. J. Structures of bacterial flagellar motors from two FliF–FliG gene fusion mutants. *J. Bacteriol.* **183**, 6404–6412 (2001).
- Sosinsky, G. E., Francis, N. R., Stallmeyer, M. J. & DeRosier, D. J. Substructure of the flagellar basal body of *Salmonella typhimurium*. *J. Mol. Biol.* **223**, 171–184 (1992).
- Stallmeyer, M. J., Hahnenberger, K. M., Sosinsky, G. E., Shapiro, L. & DeRosier, D. J. Image reconstruction of the flagellar basal body of *Caulobacter crescentus*. *J. Mol. Biol.* **205**, 511–518 (1989).
- Suzuki, H., Yonekura, K. & Namba, K. Structure of the rotor of the bacterial flagellar motor revealed by electron cryomicroscopy and single-particle image analysis. *J. Mol. Biol.* **337**, 105–113 (2004).
- Young, H. S., Dang, H., Lai, Y., DeRosier, D. J. & Khan, S. Variable symmetry in *Salmonella typhimurium* flagellar motors. *Biophys. J.* **84**, 571–577 (2003).
- Khan, S., Dapice, M. & Reese, T. S. Effects of *mot* gene expression on the structure of the flagellar motor. *J. Mol. Biol.* **202**, 575–584 (1988).
- Khan, S., Ivey, D. M. & Krulwich, T. A. Membrane ultrastructure of alkaliphilic *Bacillus* species studied by rapid-freeze electron microscopy. *J. Bacteriol.* **174**, 5123–5126 (1992).
- Khan, S., Khan, I. H. & Reese, T. S. New structural features of the flagellar base in *Salmonella typhimurium* revealed by rapid-freeze electron microscopy. *J. Bacteriol.* **173**, 2888–2896 (1991).
- Coulton, J. W. & Murray, R. G. Cell envelope associations of *Aquaspirillum serpense* flagella. *J. Bacteriol.* **136**, 1037–1049 (1978).
- Yonekura, K. *et al.* Electron cryomicroscopic visualization of PomA/B stator units of the sodium-driven flagellar motor in liposomes. *J. Mol. Biol.* **357**, 73–81 (2006).
- Braun, T. F., Al-Mawsawi, L. Q., Kojima, S. & Blair, D. F. Arrangement of core membrane segments in the MotA/MotB proton-channel complex of *Escherichia coli*. *Biochemistry* **43**, 35–45 (2004).
- Limberger, R. J. The periplasmic flagellum of spirochetes. *J. Mol. Microbiol. Biotechnol.* **7**, 30–40 (2004).
- Jackson, S. & Black, S. H. Ultrastructure of *Treponema pallidum* Nichols following lysis by physical and chemical methods. II. Axial filaments. *Arch. Mikrobiol.* **76**, 325–340 (1971).
- Katayama, E., Shiraishi, T., Oosawa, K., Baba, N. & Aizawa, S. Geometry of the flagellar motor in the cytoplasmic membrane of *Salmonella typhimurium* as determined by stereo-photogrammetry of quick-freeze deep-etch replica images. *J. Mol. Biol.* **255**, 458–475 (1996).
- Charon, N. W. & Goldstein, S. F. Genetics of motility and chemotaxis of a fascinating group of bacteria: the spirochetes. *Annu. Rev. Genet.* **36**, 47–73 (2002).
- Brown, P. N., Hill, C. P. & Blair, D. F. Crystal structure of the middle and C-terminal domains of the flagellar rotor protein FliG. *EMBO J.* **21**, 3225–3234 (2002).
- Lowder, B. J., Duyvesteyn, M. D. & Blair, D. F. FliG subunit arrangement in the flagellar rotor probed by targeted cross-linking. *J. Bacteriol.* **187**, 5640–5647 (2005).
- Leadbetter, J. R., Schmidt, T. M., Graber, J. R. & Breznak, J. A. Acetogenesis from H₂ plus CO₂ by spirochetes from termite guts. *Science* **283**, 686–689 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Matson for growing the *T. primitia* cultures. This work was supported, in part, by NIH grants to G.J.J., a DOE grant to G.J.J., a Searle Scholar Award to G.J.J., an NSF grant to J.R.L., an NIH graduate fellowship to G.E.M., and gifts to Caltech from the Ralph M. Parsons Foundation, the Agouron Institute, and the Gordon and Betty Moore Foundation.

Author Contributions G.E.M. collected and analysed the data, and drafted the text and figures; J.R.L. provided cells and discourse; and G.J.J. guided the research and manuscript editing throughout.

Author Information The averaged and symmetrized structure has been deposited in the EM Data Bank (<http://www.ebi.ac.uk/msd/index.html>) with the accession code EMD-1235. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.J.J. (jensen@caltech.edu).

Quality control

Doubt is often cast on the reliability of DNA microarrays, but resources are becoming available to help researchers overcome many of the problems inherent in this technology. Michael Eisenstein reports.

It is more than 15 years since DNA microarrays were developed, and in that time they have been adored, attacked and, in an effort to look beyond the hype, appraised. One outcome of this 'soul searching' has been the realization that the flaws inherent to gene-expression arrays are similar to those of other high-throughput platforms. "I don't think microarrays are different from other technologies in that respect, and it's important for people to keep that in mind," says Janet Warrington, vice-president of molecular diagnostics and emerging markets research and development for Affymetrix, based in Santa Clara, California. "I try to point out 'microarray exceptionalism' wherever I find it."

John Quackenbush, a computational biologist at the Harvard School of Public Health, sees several fundamental errors in the way many researchers tackle microarray gene-expression studies. "People tend to go out blindly and do experiments, then go back and try to analyse them and figure out what the question is afterwards. I think that's the first thing you have to avoid," he says. "It's also important to make sure you remove confounding factors from the experiment wherever possible."

Systematically sorting out sources of error can be a daunting process, but a recent series of investigations by people such as Quackenbush into cross-experimental, cross-platform and cross-laboratory variability between array experiments has helped clarify some issues that were preventing comparisons being made between experiments.

Several multi-institutional projects are now under way to develop more reliable experimental protocols and controls (see 'Standards and practices').

Meanwhile, many scientists, manufacturers and programmers are working to develop practical tools that could help eliminate unwanted variability from experiments and analyses.

The biggest problems often occur early on. Existing kits for RNA preparation are effective, but many are designed for a 'best-case scenario': large amounts of fresh biological source material. Many researchers are now interested in studying gene expression in a small number



Affymetrix's tiling and exon arrays: two alternatives for in-depth screening of the human genome.

of cells, necessitating efficient systems that can work with limited samples.

Various systems have been developed, many of which are based on a linear-amplification procedure known as the Eberwine method. Labelling strategies typically fall into two main categories: direct and indirect. Direct methods, used in systems such as

CyScribe, available from GE Healthcare Life Sciences of Little Chalfont, UK, and Chip-Shot, from Promega of Madison, Wisconsin, typically involve the incorporation of fluorescent-dye-conjugated nucleotides during complementary DNA (cDNA) synthesis.

One favoured indirect-labelling strategy involves incorporating an amino-alkyl-modified nucleotide into cDNA or amplified RNA transcripts, and then labelling these with

STANDARDS AND PRACTICES

Leming Shi of the US Food and Drug Administration (FDA) in Rockville, Maryland, is unabashed in his affection for microarrays. But he was disconcerted by the recent publication of several papers challenging the reliability of gene-expression microarray experiments. One article, for example, reported such disagreement that an analysis of 185 genes using three different technologies revealed concordant readings for only four transcripts (P. K. Tan *et al. Nucl. Acid Res.* 31, 5676–5684; 2003).

Subsequently, Shi and his colleagues found that an alternative analytical approach greatly improved cross-platform concordance for these data sets (L. Shi *et al. BMC Bioinformatics* 6 (Suppl. 2), S12; 2005). But the lingering climate of uncertainty, and concerns about the potentially serious implications for the use of microarray data in the FDA

drug-approval process, led them to launch the MicroArray Quality Control (MAQC) project. The MAQC brought together research leaders from government, academia and industry to establish tightly controlled 'gold standard' comparisons of microarray systems. They began by identifying commercially



Leming Shi hopes to improve the reproducibility and comparability of microarray work.

available, trustworthy 'standard' RNA samples. But this was just the start. "The MAQC's main goal is to generate a vast reference data set," says Shi. "We have conducted more than 1,000 array hybridizations with these reference samples, plus we're using three alternative technologies and we requested that each system be evaluated at three testing sites." The MAQC recently completed a review of its final data, and will present its findings in a series of articles to be published in *Nature Biotechnology* next month.

In the meantime, many in the field are awaiting the outcome of a complementary initiative: the External RNA Controls Consortium (ERCC). This evolved from a 2003 meeting headed by the US National Institute of Standards and Technology in Gaithersburg, Maryland. It aims to identify and help make commercially available a collection of reliable RNA 'spike-in'

controls, which can be included in any microarray experiment to assess variables such as labelling and hybridization efficiency. Participation has grown rapidly, and ERCC leader Janet Warrington, who is a vice-president at Affymetrix in Santa Clara, California, finds the early progress promising. "A number of organizations that were already using their own controls have donated these — no strings attached — for testing," she says. "So we have a collection of 100 to 150 controls that will be tested across platforms and we have eight sites that have volunteered to carry out the testing."

Both projects have benefited from collaborative environments that have allowed even direct competitors to work together towards a shared goal. "We all share the belief that if we're successful, we'll expand the marketplace for everyone," says Warrington.

M.E.

chemically functionalized fluorescent dyes. EPICENTRE Biotechnologies in Madison, Wisconsin, has incorporated this approach into its TargetAmp kits for RNA amplification. The company says these kits can deliver up to 5-million-fold amplification, and even allow the study of single cells. "We can get down to 10 picograms of starting RNA," says Shervin Kamkar, a technical-sales specialist at the firm, "so it's really useful for people doing stem-cell work or laser capture."

Designer labelling

Genisphere of Hatfield, Pennsylvania, uses a unique labelling approach for its 3DNA kits that is based on fluorescently tagged DNA-based dendrimers. These contain sequences complementary to 'capture' sequences added to cDNA during sample amplification. Different dendrimers are available with varying quantities of linked fluorophore molecules that determine the limits of detection, down to less than a microgram of starting material. Detection can be further augmented with the company's SenseAmp kits, which use one or two rounds of a non-Eberwine amplification strategy to produce sense-strand RNA. "We've gone as low as 0.1 nanograms of total RNA," says Bob Getts, Genisphere's director of research and development, "which is typically from ten cells."

Another problem is posed by the increasing use of microarrays for analysis of clinically prepared formalin-fixed paraffin-embedded (FFPE) samples, such as tumour biopsies. These may have been in storage for anything

from months to decades, and the resulting RNA degradation is a serious challenge for reagent designers.

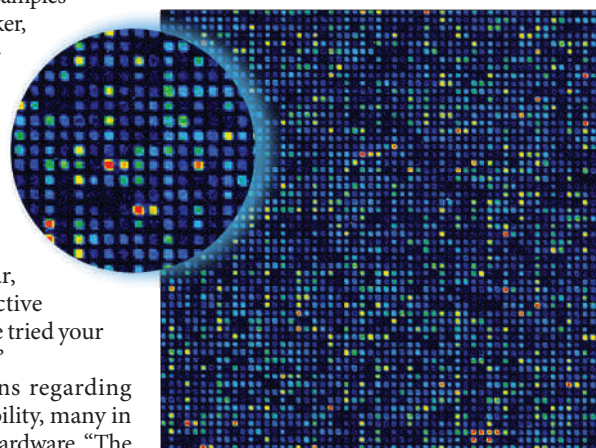
Several companies, including EPICENTRE and Genisphere, are working on this. According to Getts, SenseAmp is well-suited to FFPE work. "We routinely use samples that have been degraded to between 50 and 250 bases long," he says. Array manufacturer Illumina of San Diego, California, offers an alternative with the DNA-mediated annealing, selection, extension and ligation (DASL) assay. This is an adaptation of its GoldenGate genotyping technology that takes advantage of a universal array consisting of large numbers of specific tag sequences in order to quantify PCR-amplified primer-extension products. "It's been shown to work on samples as old as 20 years," says Shawn Baker, Illumina's scientific product manager for gene expression.

Still further optimization will be needed as researchers continue to target smaller and more biologically relevant specimens. "There's just so much variation in the material that you're given," says Kamkar, "and it's hard to know how effective a technology is until people have tried your approach on their own samples."

When asked about concerns regarding microarray experimental reliability, many in the field are quick to defend the hardware. "The microarray instrument itself, if used correctly, is precise and accurate," says Rafael Irizarry, a

biostatistician at the Johns Hopkins University in Baltimore, Maryland. "The quality of commercial arrays is improving, whereas their price is dropping, so commercial arrays are supplanting the 'home-brew' approach more and more," says Quackenbush. Different manufacturers have opted for various strategies to improve experimental quality and to minimize opportunities for human error (see 'Hands off!').

John Blume, vice-president of assay and application product development at Affymetrix, cites ever-increasing probe density and genome coverage as a secret of Affymetrix's success. Their Human Genome U133 Plus 2.0 GeneChip microarrays feature 11 different oligonucleotide probes for each transcript, which



Super-sensitive: high-density arrays of long probes are offered by NimbleGen Systems.

HANDS OFF!

To err is human. Scientists are keenly aware of this, and the multistage nature of microarray experiments provides ample opportunity for the human aspect of experimental error. Robotics could offer a solution, as well as potential for greater experimental efficiency.

"It used to be that the only people who thought about automation were the drug screeners, the

diagnostic folks, people who had to do things hundreds of thousands of times," says John Blume, vice-president of assay and application product development at Affymetrix of Santa Clara, California. "But a lot of early-stage discovery-type work now happens at the level of hundreds of things at a time."

Affymetrix uses a modular approach, developing units that automate individual experimental

stages. Several other companies have taken a similar path after finding that many customers had already taken matters into their own hands. "A lot of them have already adopted automation infrastructure," says Kevin Meldrum, director of genomics marketing at Agilent Technologies, Santa Clara, California. "They don't want to have to go out and buy a completely new system."

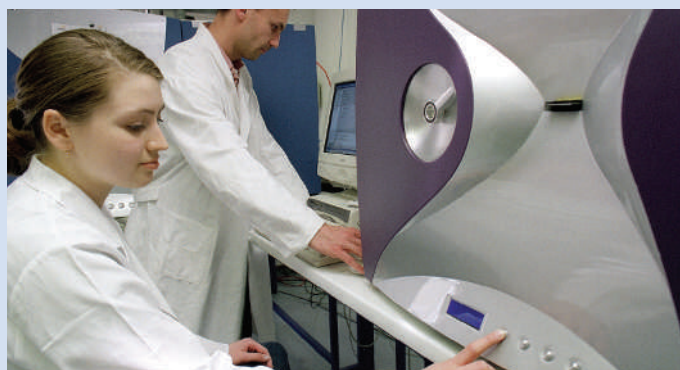
San Diego-based array manufacturer Illumina offers an 'arrays of arrays' format inherently designed for higher-throughput, and so further automation is a lower priority, although the company offers an AutoLoader that can process the scanning of up to 32 BeadChips — containing up to 256 arrays in total — in 24 hours.

Chip-builder and service provider NimbleGen Systems, based in Madison, Wisconsin, has taken an opposing approach. "Our mission is to make everything automated,"

says Emile Nuwaysir, vice-president of business development. "So that analysis is monitored and quality controlled by humans, but not run by humans." He projects near-complete automation of the full process by the end of the year.

febit Biotech in Heidelberg, Germany, brings this philosophy to the benchtop with the GENIOM, a system that fully automates most processes, including array synthesis, hybridization and analysis. Current GENIOM arrays are limited to 6,000 features but will soon expand to 15,000, and Peer Stähler, vice-president of marketing and sales at febit, believes that the instrument's speed and flexibility offer benefits for rapid experimental probe design and optimization. "You can easily import and synthesize the results of other people's bioinformatics," he says, "and you can export any capture probe that you have evaluated."

M.E.



febit's GENIOM automates many steps of microarray experiments.

confer a number of benefits. “As annotation shifts, people’s bets on which sequence is useful for a gene can prove wrong,” says Blume. “This design philosophy of multiple sequence probes per gene provides a buffer that single sequences cannot.” It also protects against unexpected glitches at the hybridization stage. In addition, Affymetrix offers broadened coverage through its genomic-tiling arrays and, more recently, new exon arrays that allow users to assemble detailed, genome-wide exon-usage profiles for human, mouse and rat studies.

Probing for answers

Illumina also uses redundancy to maintain experimental quality control in its bead-based arrays. “Each of our arrays has about 30 replicates of each probe,” says Baker. “Because these 30 measurements are spread randomly across the chip, we don’t have to worry about little things like smudges on the array — any outlier measurements get removed.” These arrays also benefit from the combination of high probe density with the inclusion of multiple arrays on a given chip. This allows users to simultaneously profile a number of samples — up to 96 parallel arrays — in an ‘array of arrays’ format.

Agilent Technologies of Santa Clara, California, touts the use of long probes — 60mers, compared with Affymetrix’s 25mers — as an advantage for enhancing sensitivity to low-abundance transcripts, typically a weakness for microarray platforms. Agilent’s instruments incorporate a multiple-scan approach, further extending the sensitivity of detection. “You can

look at a broader range of transcripts and still get linearity with regard to the signal recorded,” explains Kevin Meldrum, director of genomics marketing. Agilent has also incorporated proprietary ‘spike-in’ controls into its platform, which allow monitoring of experimental quality.

An efficient and cost-effective production process gives NimbleGen Systems of Madison, Wisconsin, particular flexibility in the generation of its arrays. These combine a maskless photolithography method with a proprietary chemical process for efficient and accurate *in situ* synthesis of high-density probe arrays. The company’s latest generation chips contain more than 2 million probes. NimbleGen also favours the use of long, typically 60mer, probes. “We are the only company that combines long oligomers with high density,” says vice-president of business development Emile Nuwaysir. NimbleGen’s rapid production process also allows it to continually update its probe sequences to align with the latest genome-annotation data. Affymetrix is currently taking advantage of this process for the production of NimbleGen-manufactured NimbleExpress custom GeneChips.

A relatively recent entrant into the gene expression array field, Applied Biosystems of Foster City, California, has used years of experience in genomic work — and access to the proprietary genome databases of Celera Genomics, based in Rockville, Maryland — to good advantage in the design of its oligonucleotide arrays. “We’ve basically front-loaded all of the bioinformatics work,” says staff scientist Chris Streck. “We do all the curation and

annotation of these particular genes, and we make sure we have the most comprehensive and complete view of the genome to begin with.” Applied Biosystems also benefits from a chemiluminescence-based approach to detection, with considerably reduced background noise relative to standard fluorescent systems.

The number crunch

However, high-quality samples and high-tech instrumentation alone won’t save the microarray experiment. Some of the most fundamental challenges lie in gleaning biological significance from mounds of data and designing experiments with a statistically sound foundation.

David Allison, a biostatistician at the University of Alabama at Birmingham, remembers the early days of microarray work with horror. “The sample sizes were way too small, unjustified statements were made, and the analyses were primitive,” he says. Fortunately, he adds, “the field recognized this, and a lot of people started weighing in with their own methods”.

According to Irizarry, an important first step for good analysis is the effective pre-processing of raw data, using algorithms that accurately convert spot fluorescence to gene-expression estimates. “Changing those algorithms can make a difference,” he says, “and you can turn an experiment that looks so-so into something that looks powerful and precise.” Irizarry has also called attention to the importance of data normalization, and designed an online tool, Affycomp II, which allows users to benchmark their normalization methods using ‘known’

SHARE AND SHARE ALIKE

Many working with microarrays now recognize that one way uncertainty about experimental findings can be dispelled is by being more transparent about methodology and data. This realization has transformed the field. For instance, after some initial resistance, almost every major commercial vendor has made the sequences and annotations of their probes publicly available — to the considerable benefit of the community as a whole.

This awareness has also manifested itself in the drive to develop shared resources for pooling experimental data and systems for clearly defining how these data were obtained. A leading force in this regard is the Microarray Gene Expression Data (MGED) Society, which put forward a proposal in 2001 for experimental annotation standards known as minimum information about a microarray experiment (MIAME),

designed to record key details about factors such as sample preparation and experimental design. These standards were embraced by many, and several leading journals, including *Cell*, *The Lancet* and *Nature*, demand MIAME compliance from all microarray research submissions. However, some aspects of MIAME have proved problematic.

“I think almost all academic biologists embrace the concept of openly sharing data,” says Catherine Ball of Stanford University in California, the current president of the MGED Society. “But embracing the process and actually taking part are very different, and it can be difficult to fully annotate your data.” According to Gavin Sherlock, also of Stanford and MGED, part of the problem was MAGE-ML (microarray and gene expression markup language), the XML-based language initially developed for

MIAME data recording. “Nobody can look at it, nobody can read it, nobody can edit it,” he says. “It’s very difficult to use.” This is reflected in the uploading of data to public databases, another process strongly advocated by MGED.

The ArrayExpress database of the European Bioinformatics Institute in Cambridge, UK, is strictly

MIAME-compliant, and receives considerably fewer submissions than the non-MIAME-compliant Gene Expression Omnibus (GEO) of the National Center for Biotechnology Information in Rockville, Maryland. MGED is now poised to release a considerably simpler format for data submission, and Ball is hopeful that this, along with other user-friendly software tools, will make a difference.

But, fundamentally, compliance comes down to the effort scientists can and will put in. All of the MicroArray Quality Control project’s data are being deposited into both GEO and ArrayExpress, and although this has proved an onerous task, Leming Shi of the US Food and Drug Administration sees clear rewards in the effort. “Depositing the data may be a painful process, but we have to do it for the sake of the community,” he says. “The more information we have in the future, the better.” M.E.



Catherine Ball believes simpler software tools could encourage better MIAME compliance.

COMPANY	PRODUCTS/ACTIVITY	LOCATION	URL
Catalogue and custom arrays			
Affymetrix	GeneChip oligonucleotide arrays, array processing instruments and analytical software	Santa Clara, California	www.affymetrix.com
Agilent Technologies	Oligonucleotide arrays, processing hardware and software	Santa Clara, California	www.agilent.com
Applied Biosystems	Oligonucleotide arrays, chemiluminescent detection tools, Ambion reagents for sample preparation and labelling	Foster City, California	www.appliedbiosystems.com
Atactic Technologies	Speciality custom microarrays	Houston, Texas	www.atactictech.com
Aviva Systems Biology	ChIP-GLAS arrays for chromatin immunoprecipitation	San Diego, California	www.avivasysbio.com
BioCat	Premade and custom arrays and reagents	Heidelberg, Germany	www.biocat.de
BioWindow	Catalogue and custom expression arrays and array services	Shanghai, China	www.biowindow.com
CapitalBio Corporation	Premade expression profiling arrays	Beijing, China	www.capitalbio.com
CLONDIAG	ArrayTube biochip format for diagnostic applications	Jena, Germany	www.clondia.com
Clontech (Takara Bio)	Atlas arrays and products for performing microarray experiments	Mountain View, California	www.clontech.com
Combimatrix	Custom and catalogue arrays, CustomArray synthesizer	Mukilteo, Washington	www.combimatrix.com
Eppendorf	DualChip DNA microarrays	Hamburg, Germany	www.eppendorf.com
Eurogentec	Probe sets for array use, ArrayIt printing tools, array design services	Seraing, Belgium	www.eurogentec.com
GE Healthcare	CodeLink oligonucleotide arrays and processing tools	Little Chalfont, UK	www.amershambiosciences.com
GenoSensor Corporation	GenoExplorer Biochips for expression analysis, array services	Tempe, Arizona	www.genosensorcorp.com
Illumina	BeadArray platform for oligonucleotide arrays and associated instruments and reagents	San Diego, California	www.illumina.com
Integrated DNA Technologies	Custom oligonucleotides for use on arrays	Coralville, Iowa	www.idtdna.com
Microarrays	Catalogue arrays and custom array printing services	Nashville, Tennessee	www.microarrays.com
Miltenyi Biotec	PIQOR microarrays, sample preparation reagents, analytical services	Bergisch-Gladbach, Germany	www.miltenyibiotec.com
Nimblegen	Custom and catalogue array designs, array experimental services	Madison, Wisconsin	www.nimblegen.com
Ocimum Biosolutions	Oligonucleotide-based OciChips and hybridization services	Indianapolis, Indiana	www.ocimumbio.com
Operon Biotechnologies	Pre-printed oligonucleotide arrays and probe sets for microarray use	Huntsville, Alabama	www.operon.com
Orion Genomics	MethylScope microarrays for analysing genomic methylation states	St Louis, Missouri	www.oriongenomics.com
Oxford Gene Technology	Oligonucleotide arrays for a variety of applications	Oxford, UK	www.ogt.co.uk
Panomics	cDNA arrays and associated reagents	Fremont, California	www.panomics.com
Phalanx Biotech	Human OneArray genomic array and hybridization kits	Hsinchu, Taiwan	www.phalanxbiotech.com
SuperArray	GEArray preprinted arrays, hybridization and analysis services	Frederick, Maryland	www.superarray.com
Telechem International	ArrayIt oligonucleotide microarrays, scanners and instruments	Sunnyvale, California	www.arrayit.com
Zeptosens	Custom oligonucleotide arrays	Witterswil, Switzerland	www.zeptosens.com
Systems for array production			
ArrayJet	Inkjet-based microarray printers	Mayfield, Scotland	www.arrayjet.co.uk
BioAutomation	MagnaSpotter for pin-based microarray printing	Plano, Texas	www.bioautomation.com
BioForce Nanosciences	NanoArrayer for desktop array production	Ames, Iowa	www.bioforcenano.com
Bio-Rad Laboratories	Arraying systems, hybridization chambers, software tools	Hercules, California	www.bio-rad.com
febit	GENIOM platform for array design, synthesis and hybridization	Heidelberg, Germany	www.febit.de
Genetix	Pin-based arrayers and array scanners	New Milton, UK	www.genetix.com
Genomic Solutions	OmniGrid microarrayers and hybridization chambers	Ann Arbor, Michigan	www.genomicsolutions.com
GeSiM	NanoPlotter for non-contact microarraying	Großkrummannsdorf, Germany	www.gesim.de
Parallel Synthesis Technologies	Silicon Microarray technology for array printing	Santa Clara, California	www.parallel-synthesis.com
PerkinElmer	Contact and non-contact array spotting, array scanners, software	Wellesley, Massachusetts	las.perkinelmer.com
Plexigen	Array printing and reading for GeneCube platform	Cary, North Carolina	www.plexigen.com
V&P Scientific	Glass-slide and compact-disc microarrayers	San Diego, California	www.vp-scientific.com
Xenopore	Manual microarrayers and slide scanning services	Hawthorne, New Jersey	www.xenopore.com
Array scanners			
Alpha Innotech	Array scanners	San Leandro, California	www.alphainnotech.com
Applied Precision	arrayWoRx biochip reader	Issaquah, Washington	www.api.com
Biomedical Photometrics	Microarray scanners and analytical software	Waterloo, Canada	www.confocal.com
Biomolex	Microarray readers based on isotopic detection	Oslo, Norway	www.biomolex.com
Fujifilm	Scanners and analytical software	Tokyo, Japan	www.fujifilm.com
LaVision BioTec	BioAnalyzer CCD-based platforms for microarray imaging	Bielefeld, Germany	www.lavisionbiotec.de
MiraiBio	FMBIO fluorescent scanner	Alameda, California	www.mirai.bio.com
Molecular Devices	Microarray scanners and analytical software	Sunnyvale, California	www.moleculardevices.com
Tecan	Microarray scanners and hybridization stations	Zurich, Switzerland	www.tecan.com
VIDAR Systems	Revolution 4200 microarray scanner	Herndon, Virginia	www.microarrayscanner.com
Reagents for sample preparation			
Active Motif	Reagents for RNA preparation	Carlsbad, California	www.activemotif.com
Agencourt Bioscience	SPRI reagents for nucleic-acid purification	Beverly, Massachusetts	www.agencourt.com
Ambion	Kits for RNA isolation, labelling and amplification	Austin, Texas	www.ambion.com
Corning Life Sciences	Microarray reagents and slides	Acton, Massachusetts	www.corning.com
Enzo Life Sciences	Kits for labelling and amplification	Farmingdale, New York	www.enzolifesciences.com
Episcentre Biotechnologies	Kits for sample preparation	Madison, Wisconsin	www.epibio.com
Genisphere	Kits for RNA amplification and dendrimer-based sample labelling	Hatfield, Pennsylvania	www.genisphere.com
Invitrogen	BioModule kits for sample preparation	Carlsbad, California	www.invitrogen.com
Kreatech Biotechnology	Labelling kits for gene-expression analysis	Amsterdam, the Netherlands	www.kreatech.com
NuGEN Technologies	Ovation RNA amplification and labelling systems	San Carlos, California	www.nugenttechnologies.com

COMPANY	PRODUCTS/ACTIVITY	LOCATION	URL
Promega	Pronto! Plus system for sample isolation and labelling	Madison, Wisconsin	www.promega.com
QIAGEN	Array-ready oligo sets and kits for sample preparation	Hilden, Germany	www.qiagen.com
Roche Applied Sciences	Kits for RNA isolation, labelling and amplification	Indianapolis, Indiana	www.roche-applied-science.com
Sigma-Aldrich	Kits for sample preparation	St Louis, Missouri	www.sigmaaldrich.com
Stratagene	Labelling kits, RNA controls, analytical software	La Jolla, California	www.stratagene.com
Software for microarray analysis			
Applied Maths	GeneMaths XT software for microarray analysis	Sint-Martens-Latem, Belgium	www.applied-maths.com
Ariadne Genomics	PathwayStudio and PathwayExpert software for pathway assembly	Rockville, Maryland	www.ariadnegenomics.com
BioAware	BioArchitech software for the design of microarrays	Hannut, Belgium	www.bio-aware.com
BioBase	Regulatory and signal-transduction databases	Wolfenbüttel, Germany	www.biobase-international.com
BioDiscovery	Software programs for experimental analysis and data management	El Segundo, California	www.biodecovery.com
Biomind	ArrayGenius software for data analysis	Rockville, Maryland	www.biomind.com
Biotique Systems	BLIS software for data integration	Reno, Nevada	www.biotiquesystems.com
BlueGnome	BlueFuse software for analysis of expression and CGH arrays	Great Shelford, UK	www.cambridgebluegnome.com
DNASTAR	ArrayStar data-visualization software	Madison, Wisconsin	www.dnastar.com
Genedata	Expressionist software for microarray analysis	Basel, Switzerland	www.genedata.com
GeneGo	MetaCore software for microarray analysis	St Joseph, Michigan	www.genego.com
Genesifter	Software for microarray data management	Seattle, Washington	www.genesifter.net
Genomatix	ChipInspector and BiblioSphere software	Munich, Germany	www.genomatix.de
Golden Helix	HelixTree software for SNP analysis	Bozeman, Montana	www.goldenhelix.com
Improved Outcomes	GeneLinker software for gene-expression analysis	Kingston, Canada	www.improvedoutcomes.com
Ingenuity Systems	Pathways Analysis software for building networks from microarray data	Redwood City, California	www.ingenuity.com
Insightful	S+ArrayAnalyzer software for statistical analysis	Seattle, Washington	www.insightful.com
Partek	Genomics Suite for analysis of microarray data	St Louis, Missouri	www.partek.com
PREMIER Biosoft International	Array Designer software	Palo Alto, California	www.premierbiosoft.com
Rosetta Biosoftware	Rosetta Resolver for pathway analysis	Seattle, Washington	www.rosettabio.com
SilicoCyte	Software for microarray annotation and analysis	Chesterfield, Missouri	www.silicocyte.com
Spotfire	DecisionSite for Microarray Analysis software	Somerville, Massachusetts	www.spotfire.com
Strand Life Sciences	Multiple software products for microarray design and analysis	Bangalore, India	www.strandgenomics.com
Vigene Tech	Microarray image analysis software	Carlisle, Massachusetts	www.vigenetech.com

Microarray service providers

AROS Applied Biotechnology	Sample preparation and expression analysis services	Aarhus, Denmark	www.arosab.com
Asuragen	Expression profiling services	Austin, Texas	www.asuragenservices.com
Codon Biosciences	Various microarray services	Houston, Texas	www.codonbiosciences.com
Cogenics	Standard and custom expression array services	Newton, Massachusetts	www.cogenics.com
CTL BIO Services	Gene-expression array services	Rockville, Maryland	www.ctlbio.com
Expression Analysis	Sample purification and expression array services	Durham, North Carolina	www.expressionanalysis.com
Gene Logic	Microarray data generation and analysis services	Gaithersburg, Maryland	www.genelogic.com
Geneservice	Gene-expression microarray services	Cambridge, UK	www.geneservice.co.uk
Genome Explorations	Microarray hybridization and analysis services	Memphis, Tennessee	www.genome-explorations.com
Genotypic Technology	Gene-expression microarray services and analytical software	Bangalore, India	www.genotypictech.com
Integrated Genomics	Microarray design, manufacturing and analysis services	Chicago, Illinois	www.integratedgenomics.com
Ipsogen	Gene-expression profiling with an emphasis on cancer diagnostics	Marseille, France	www.ipsogen.com
LC Sciences	Various microarray services	Houston, Texas	www.lcsciences.com
MOgene	Gene-expression microarray services	St Louis, Missouri	www.mogene.com
Precision Biomarker Resources	Hybridization and analysis services for expression, exon and tiling arrays	Evanston, Illinois	www.precisionbiomarker.com
Quark Biotech	Gene-expression microarray services	Fremont, California	www.quarklabs.com
VBC-GENOMICS	GeneChip microarray services, slides for array preparation	Vienna, Austria	www.vbc-genomics.com

Other microarray-related products

Advantix	ArrayBooster and SlideBooster for improving hybridization	Brunnthal, Germany	www.advantix.de
BioMicro Systems	MAUI microarray hybridization chambers	Salt Lake City, Utah	www.biomicro.com
Bioslide Technologies	Microarray slides	Walnut, California	www.bioslide.com
Erie Microarray	Slides and treatments for microarray preparation	Portsmouth, New Hampshire	www.eriemicroarray.com
Full Moon BioSystems	Microarray slides and QC kits, array printing services	Sunnyvale, California	www.fullmoonbio.com
Metrigenix	Proprietary Flow-thru Chip array platform, array analysis services	Toronto, Canada	www.metrigenix.com
Nanogen	NanoChip Electronic Microarray platform for genomic analysis	San Diego, California	www.nanogen.com
QUANTIFOIL Instruments	Microarray buffers and reagents	Jena, Germany	www.qinstruments.de
SCHOTT Nexterion	Slides, oligonucleotide sets, and microarray reagents	Louisville, Kentucky	www.us.schott.com/nexterion
SciGene	Array processing systems	Sunnyvale, California	www.scigene.com
Tm Bioscience	Reagents for microarray experiments	Toronto, Canada	www.tmbioscience.com

● see advertisement

naturejobs

**THE CAREERS
MAGAZINE FOR
SCIENTISTS**

A recent story in satirical newspaper *The Onion* summed up the problems faced by women today. The piece, headlined “National Organization for Women Turns 39 Again”, sought to tweak stereotypes about women. But in the end, it embraced them — and served to illustrate the prevalence of retrograde attitudes. The topic is as live in science as in any other sphere, following the departure in July of Larry Summers as president of Harvard University in the wake of comments he made about innate differences between men and women.

These comments were addressed head on in a recent Commentary by Ben Barres (*Nature* **442**, 133–136; 2006) — an article that generated a lot of comment, some of which underlined uncomfortable truths about the representation of women in science. One correspondent, for example, was prompted to review the awards given out by the UK Biochemical Society and discovered that only 3.2% of the society’s prizes have been given to women (A. C. Dolphin *Nature* **442**, 868; 2006).

Elsewhere, a report by the InterAcademy Council revealed that women typically make up less than 5% of the membership of national science academies. And a close look at more than 4,000 life scientists over 30 years has shown that women secure patents at a much slower rate than their male colleagues (see page 973).

Despite such dismal statistics, there are signs that things are changing. One scheme at the US National Institutes of Health yielded no grants to women in its first year, but after a revamp it awarded 43% to women the following year (M. Carnes *Nature* **442**, 868; 2006). The list of scientific prizes aimed specifically at women is also growing, including one from L’Oréal and a new award for South African women. But schemes to give money to women should be eyed with caution: true progress can only be claimed when women are proportionately represented in terms of conventional grants and awards. Otherwise, the irony of a quote from the fictitious president of *The Onion*’s women’s group will continue to hit too close to home: “There’s no reason why we can’t continue to make amazing achievements in our 39th year and in all our other subsequent 39th years.”

Paul Smaglik, *Naturejobs* editor

CONTACTS

Publisher: Ben Crowe
Editor: Paul Smaglik
Assistant Editor: Gene Russo

US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager/Corporations:
Peter Bless
Classified Sales Representatives
Tel: +1 800 989 7718
East USA/Canada: Andrew Bennie
**NIH/Maryland/New York/
Pennsylvania:** Shelley Cohen

San Francisco Office
Classified Sales Representative:

Michaela Bjorkman
West USA/West Corp. Canada
225 Bush Street, Suite 1453
San Francisco, CA 94104
Tel: +1 415 781 3803
Fax: +1 415 781 3805
e-mail: m.bjorkman@naturesf.com

European Head Office, London
The Macmillan Building,
4 Crinan Street, London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
Advertising Production Manager:
Stephen Russell
To send materials use London
address above.
Tel: +44 (0) 20 7843 4816

Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com
Naturejobs web development:
Tom Hancock
Naturejobs online production:
Catherine Alexander

European Satellite Office
Patrick Phelan
e-mail: p.phelan@nature.com

Japan Head Office, Tokyo
Chiyoda Building,
2-37 Ichigayatamachi,
Shinjuku-ku,
Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe
e-mail: a.watanabe@natureasia.com

The key

The future's in your hands.

Ian Whates

It's amazing what a bunch of keys can say about a person. Key-rings and their contents hold hidden depths, or so Carl had always maintained.

Take his wife's, for example: keys for the front door, car, garage and a Yale for her mother's... plus various superfluous attachments: a pink plastic pig, a Perspex heart displaying the pseudo-word 'whateva', and a smiley-emblazoned disc designed to impersonate a coin when liberating supermarket trolleys.

His own set was far more practical. Keys for car and home, one for a suitcase and another for a young lady's flat that he trusted his wife would never notice or question. Two add-ons: a worn leather fob from his very first jalopy and a pizze-shaped plait of woven leather that he'd been assured was a fertility symbol but probably wasn't.

Then there was the set he had recently 'acquired'. Six keys plus three attachments: a circular Mercedes emblem, matching one of the keys; a tiny plastic-encased photo of a girl's face — presumably the owner's daughter; and a small, squat figurine with blood-red crystal eyes. This last attachment vaguely resembled an owl and gave Carl the creeps. Quite what it said about the owner he preferred not to dwell on.

It was the keys that really intrigued him. Two differently cut front-door keys, suggesting two homes, Merc and Land Rover keys — a car for each dwelling — and two others less easily identified.

Sammy-the-Locksmith's considered opinion proved as much use as a chocolate teapot. "One's for a cabinet and the other a safety-deposit box."

"Any way of telling where it is?"

"Nope."

He wouldn't have cared, except for the small matter of the reward. A ridiculous sum, offered for the keys' return with no mention of the Gucci wallet lifted at the same time, nor of the cash and credit cards contained therein. One of these keys was clearly important to someone and therefore valuable. Carl knew which his money would've been on. It remained useless to him, however, unless he could find precisely what it opened. To his growing frustration, unlocking that particular enigma proved beyond him.

"Mightn't even be in this country," his best and final hope had concluded with a shrug.

Reluctantly he arranged a meet, at a time and place of his choosing: a bar where he was known and felt safe. His recent victim and prospective benefactor awaited — a tall, muscle-broad individual who, even in an Armani suit, failed to look entirely polished or civilized. The rugged edges were still there: an uncut gem in a presentation box.

Carl would have preferred a dead-drop, an exchange without ever meeting face-to-face, but the other would have none of it. So he watched the man arrive from across the street, alert for any hint of police or other presence. Seeing none, he entered, glancing at the barman, whose shake of the head still fell short of total reassurance.

He took a deep breath and committed himself by sitting down. Eyes locked across a table. Unwavering self-confidence and steely strength couched within grey-blue irises; this was not a man to trifle with.

"You have the keys?" The voice was relaxed and casual to the point of being unnerving.

"You got the money?"

An envelope, produced from a pocket and then slid across the table. A fat envelope.

Carl reached out but the other's hand clung to its far edge. "The keys first."

"Not until I've counted it."

A frozen tableau that persisted for time-stretching seconds until the man abruptly let go. Carl opened the envelope and flicked through the wad of fifties, not counting with any accuracy, just checking.

Satisfied, he nodded to the barman, who left his station and came across with the keys.

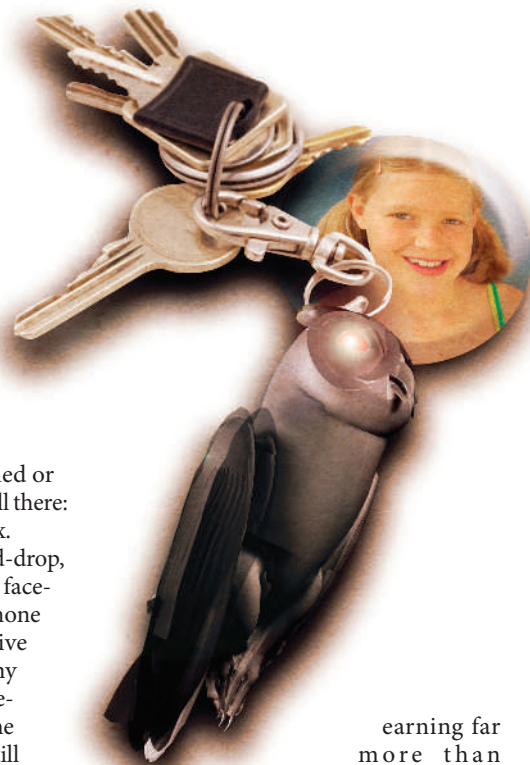
To his credit, the stranger guffawed and nodded appreciation at such complicity. He looked the keys over once before pocketing them and rising to his feet. There he paused, fixing Carl with a glare — the first suggestion of either anger or menace.

"Don't cross my path again."

"Just a minute," Carl blurted out as the man turned to leave. "You've got them back now, so you can tell me, why are they so important?"

The man smiled — a malicious, satisfied expression, which lacked any hint of humour. "Do you really imagine I'm going to tell you?"

Carl watched the retreating back until it was out the door and away. Despite



earning far more than anticipated from this episode he still felt cheated, as if opportunity had somehow slipped through his grasp. What wealth or secrets had the key represented? Too late now. He would never know.

'Never' lasted a month.

Carl was watching the news. He remained sceptical and unmoved by the inescapable buzz about the first proof of ET, scoffing at the media frenzy and avoiding television's blanket coverage — until now. He stared in disbelief at the image of what was allegedly an alien artefact. Set against a neutral background, the picture provided no sense of proportion, but Carl knew at once that it was small: a tiny, squat, owl-like effigy with blood-red eyes.

The reporter — all blonde hair, glossy lipstick and gushing exuberance — was explaining how its eccentric owner and discoverer had little faith in conventional secure repositories, so kept the priceless item on his key-ring while awaiting the vital test results, which were soon to confirm its non-terrestrial composition. Thus inspired, she dusted down her poetic licence and waxed lyrical about the artefact's potential for unlocking new worlds, labelling it 'The Key to the Future'.

Carl switched off the TV. For long minutes he sat there, simply staring at the empty screen.

■
Having recently sold stories to various websites, magazines and a book anthology, Ian Whates is currently editing and publishing a limited-edition chapbook with new stories from Stephen Baxter, Ian Watson, Liz Williams, Jon Courtenay Grimwood and others.

Abstractions



FIRST AUTHOR

Alicia Soderberg, an astronomy student at the California Institute of Technology, bet her PhD on capturing brilliant, long bursts of light from a relatively close dying star. Such bursts are expected from only a rare type of supernova — the gamma-ray burst (GRB). By studying local supernovae, Soderberg and her team aimed to pinpoint the physical mechanism distinguishing an ordinary supernova from a GRB or its X-ray flash (XRF) subclass. The tell-tale signal eventually came from the second-nearest GRB ever recorded (see page 1014). Soderberg talked to *Nature* about her results.

How did you catch the bursts?

We coordinated our efforts across many of the most powerful observatories in the Northern and Southern Hemispheres. We also used simultaneous observations from several satellites, including the Chandra X-ray Observatory in Massachusetts and NASA's Swift X-ray Telescope. The 'afterglow' light from GRBs and XRFs fades quickly, so rapid, coordinated responses are essential. When the Swift satellite detects a new burst, it rings our cellphones, so we can immediately begin coordinating our observational response.

What sorts of things were you doing when the satellite rang you?

Mostly sleeping! The Swift satellite tends to find bursts at night, so optical telescopes can begin observing the afterglow immediately.

How long have you been working on this?

I've spent four to five years observing new supernovae, looking for the tell-tale sign of an accompanying GRB or XRF.

What was your reaction to the data?

I was excited and relieved! When I began my thesis I wondered whether GRBs and XRFs might be lurking in a significant fraction of local supernovae. But after studying about 150 supernovae and finding no sign of a GRB or XRF, I wondered if I would ever find one.

How hard was it piecing together data from so many instruments?

It can be very confusing. Often the data don't fit the classical models.

What drew you to this field in the first place?

The GRB field is young, so every event brings new discoveries.

What's the most significant thing you learned?

Supernovae and GRBs might not be as different as was believed. We now know that there are explosions that bridge the gap between ordinary supernovae and the powerful GRBs and XRFs. ■

MAKING THE PAPER

Martin Cohn

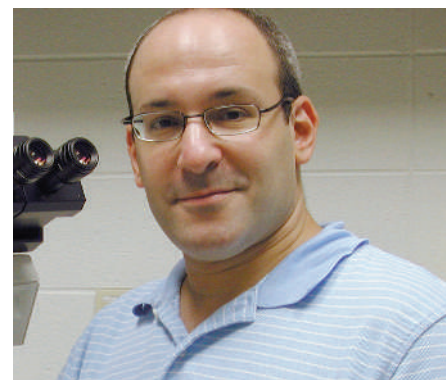
Fish fossils and sharks offer clues to the origin of vertebrate limbs.

Chinese fossil finds and the lamprey genome both provided insight into a quest to understand the evolution of fins. In the course of their research, developmental biologist Martin Cohn, based at the University of Florida, and his graduate students Renata Freitas and GuangJun Zhang even attempted to fluorescently label shark embryos. Their efforts identified a surprise 'recycling' of genetic mechanisms involved in the formation of median and, much later in evolution, paired fins (see page 1033).

Before this, most fin-development work had focused on the zebrafish. But Cohn and his colleagues wanted to understand the gene-expression patterns that give rise to paired fins, so they turned to the shark, the most primitive living vertebrate with such fins. During their initial gene-expression screen, they noticed that genes expressed in paired fins are also active in median fins.

While they were redirecting their efforts to study median fins, an exciting find of fossil fishes in China confirmed that median fins arose about 100 million years before paired fins. This suggested that median fins might hold the key to the origin of fins and limbs, and, therefore, the first steps towards vertebrate locomotion.

Those interested in locomotion had paid little attention to median fins, but one previous study of zebrafish had determined that neural-crest cells — embryonic cells that give rise to cranial skeletal and connective tissues — are involved in median fin development. "That one experiment gave rise to the idea that neural-crest cells form the median fins," says Cohn. To determine whether neural-crest cells and/or somitic cells — embryonic cells that develop into muscles and vertebrae — have a role, the researchers first attempted to microinject shark eggs with a fluorescent label.



But working with unusual animal models often introduces time-delaying technical hurdles. Sharks are seasonal breeders, which limits the availability of embryos. And, as Cohn and his colleagues found, embryos could not survive the saltwater intrusion that occurred during introduction of the fluorescent label.

So they pursued a molecular approach, cloning various genes that mark distinct cell types to determine which give rise to median fins. They found that although neural-crest cells do make a contribution, this is minor; somitic cells form the bulk of the fin. Cohn was most surprised to find that the genetic programme for fin development was first assembled in the somitic cells, then later in evolution redeployed to the tissue that gives rise to paired fins and limbs.

The scientists went on to document gene expression in lampreys, eel-like organisms with median fins but no paired fins. The shark and lamprey straddle the origin of paired-fin nodes along the evolutionary time scale. The lamprey's median fins also proved to originate from somitic cells.

The next question for Cohn's group is 'how old is fin development?' To find out, they'll study median fin development in amphioxus, the closest living invertebrate relative of vertebrates. "What's been most exciting for us is the flow of information between developmental biology and palaeobiology," says Cohn, adding that the strength of evolution and development research is the integration of a wide range of disciplines in an effort to address evolutionary questions. ■

KEY COLLABORATIONS

Finding and analysing gamma-ray bursts requires cooperation, but also creates competition, as shown by four papers in this issue (see pages 1008–1020). All of these groups started at the same point — with a signal from NASA's Swift satellite, which reports pulses of energy from dying stars within the cosmos.

Various groups scrambled to get in touch with someone at an observatory in an ideal position to measure the waves

emitted from the bursts. Many teams shared instruments, such as the Very Large Array Telescope in New Mexico, or teamed up with observers who were sometimes on the other side of the world.

Some groups had access to their own ground-based telescopes. As a result of these alliances, the effort yielded overlap among some of the 107 authors involved. Three authors appear on two of the papers, representing a collaboration

between Pennsylvania State University, the Goddard Space Center in Greenbelt, Maryland, and the California Institute of Technology.

The other two papers share seven authors as a result of collaborations between the Max Planck Institute for Astrophysics in Germany, the University of Tokyo in Japan, the University of California, Berkeley, and the National Institute of Astrophysics in Trieste, Italy. ■